

UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR



TRABAJO FIN DE MÁSTER

Cambios epigenómicos y transcriptómicos durante el envejecimiento en pez killi

Máster Universitario en Bioinformática y Biología Computacional

Autor: DE LOS REYES BENITEZ, Alejandro

Tutor: TENA AGUILAR, Juan Jesús

Departamento de Epigenómica en enfermedades y envejecimiento

FECHA: Junio, 2021

INDICE	1
1 RESUMEN	2
2 PALABRAS CLAVE	2
3 ABSTRACT	2
4 KEY WORDS	2
5 ABREVIATURAS	3
6 INTRODUCCION	4
7 RESULTADOS Y DISCUSION	8
7.1 Análisis del transcriptoma	8
7.2 Análisis de elementos reguladores.....	13
7.3 Integración de los datos de RNA y ATAC-seq	20
7.4 Genes con dinámicas específicas de tejido	23
7.5 Genes de desarrollo en envejecimiento.....	24
8 CONCLUSION.....	27
9 METODOS.....	28
9.1 Diseño experimental	28
9.2 Genes con expresión diferencial.....	28
9.3 Regiones de la cromatina con accesibilidad diferencial	29
9.4 Integración RNA-seq y ATAC-seq.....	31
9.5 Identificación de genes específicos de tejido	32
9.6 Enriquecimiento en términos GO	32
9.7 Identificación de ortólogos.....	32
9.8 Análisis de asociación con el envejecimiento	33
9.9 Inyección de embriones.....	33
10 BIBLIOGRAFÍA.....	39

RESUMEN

El envejecimiento es un proceso universal que supone el declive de las funciones fisiológicas de los organismos. Es factor de riesgo para diversas enfermedades y su última consecuencia es la muerte. Al ser el cerebro un regulador clave de la homeostasis en organismos multicelulares, los efectos del envejecimiento en este tejido pueden afectar a otros procesos. Para el estudio del envejecimiento se han empleado organismos modelo invertebrados debido a la relativamente alta longevidad de organismos vertebrados. En los últimos años, ha aparecido un nuevo modelo vertebrado de vida corta que ha ganado interés en este campo, el pez killi (*Nothobranchius furzeri*). En este trabajo se ha estudiado cómo el envejecimiento es regulado en el cerebro del pez killi atendiendo a la información transcriptómica y epigenómica. Para ello se ha analizado genes y regiones accesibles de la cromatina usando las técnicas RNA-seq y ATAC-seq. Nuestros resultados sugieren que redes regulatorias relacionadas con el desarrollo aparecen alteradas durante el envejecimiento. Esto aporta evidencias en favor de la teoría del antagonismo pleiotrópico, donde se defiende que genes beneficiosos durante las primeras etapas de la vida pueden ser perjudiciales en etapas tardías.

PALABRAS CLAVE

Cerebro, envejecimiento, epigenética, pez killi, transcriptómica

ABSTRACT

Aging is an universal process involving the decline of the physiological function in organisms. Also, it is a risk factor for different diseases and its last consequence is death. The brain is a central modulator of homeostasis in multicellular organism, so aging effects on this tissue can influence other processes in the body. With the purpose of studying aging, invertebrate model organisms have been used due to the relatively long lifespan of vertebrate model organisms. In recent years, a new short-lived vertebrate organism has emerged as a promising model for the study of aging, the killifish (*Nothobranchius furzeri*). In this work, the regulation of aging has been studied in the killifish brain through the analysis of transcriptomic and epigenomics data. For that, gene expression and accesible chromatin regions have been analysed performing RNA-seq and ATAC-seq assays. Our results suggest that gene regulatory networks involving development are altered during aging. This adds evidence in favor of the antagonistic pleiotropy theory, which states that beneficial genes in early stages can cause functional decline later in life.

KEY WORDS

Aging, brain, epigenetics, killifish, transcriptomics

ABREVIATURAS

RNA-seq, RNA sequencing; ATAC-seq, Assay of Transposase Accessible Chromatin sequencing; KO, knockout; bp, base pair; kb, kilo bases

INTRODUCCION

El envejecimiento se define como el declive temporal de las funciones fisiológicas necesarias para la supervivencia del organismo cuya última consecuencia es la muerte. Además, existe una estrecha relación entre el paso del tiempo en los organismos y el aumento de la susceptibilidad a enfermedades comunes, como son enfermedades renales, cardiovasculares, artritis o enfermedades neurodegenerativas (Melzer et al., 2020). La prevalencia de efectos y enfermedades asociadas al envejecimiento son cada vez más comunes. El aumento de la esperanza de vida experimentado por la sociedad en las últimas décadas es responsable de ello. El porcentaje de población que supera los 60 años de edad era del 11% en 2013 y se estima su incremento hasta un 22% para 2050 (Newgard & Sharpless, 2013). A pesar de que los efectos del envejecimiento se manifiestan en diferentes tejidos, el cerebro es especialmente sensible a este proceso, dada su labor esencial en el control de la homeostasis del organismo. Es por eso que existe especial interés en el estudio del cerebro durante el proceso de envejecimiento.

En el cerebro, el declive causado por la edad se manifiesta en la disminución de las capacidades cognitivas, percepción sensorial y coordinación motora, afectando a todos los individuos en diferente grado de severidad. Mientras que algunos individuos desarrollan un envejecimiento sano otros sufren desórdenes neurológicos, siendo la enfermedad de Alzheimer y la enfermedad de Parkinson las más comunes. A nivel celular y molecular, muchos marcadores han sido descritos en el cerebro de los mamíferos, incluyendo disfunción mitocondrial, acumulación intracelular de daño oxidativo, metabolismo energético desregulado, sistemas de autofagia-lisosoma y proteasoma desapareados, incorrecta señalización en la respuesta a estrés, capacidad disminuida de reparación del ADN, actividad de la red neural anormal, gestión de los niveles de Ca^{2+} desregulado, agotamiento de células madre e inflamación. Debido a que las rutas metabólicas y sistemas reguladores tiene un alto grado de interdependencia, ninguno de esos marcadores ocurre de forma aislada (Mattson & Arumugam, 2018).

Con el fin de comprender e integrar todos los rasgos del envejecimiento han aparecido diversas teorías sobre las causas del envejecimiento. Tradicionalmente ha existido un gran apoyo a las teorías que defendían que la acumulación de daño es un proceso espontáneo que sigue las leyes de entropía. Hecho que se ve apoyado por su coherencia con la teoría de la evolución, debido a la ausencia de selección natural en el período post-reproductivo (da Costa et al., 2016). Sin embargo, en las últimas décadas se han hecho importantes descubrimientos, no sólo relacionados con comprender los mecanismos del envejecimiento sino con el desarrollo de intervenciones que extienden la longevidad y alargan el estado saludable de los órganos y sistemas fisiológicos en las últimas etapas de la vida. Ya existen evidencias que demuestran que el envejecimiento es un proceso plástico que se ve afectado por diversos factores genéticos y ambientales y que es regulado por rutas de señalización clásicas y factores de transcripción. Entre los mecanismos más estudiados se encuentra el eje que regula la detección de nutrientes y respuesta al estrés, pudiendo ser alterado con dietas deficientes/abundantes

en calorías. Cuando el alimento es abundante, los programas genéticos de las células promueven el crecimiento y la reproducción, mientras que ante condiciones adversas con alimento insuficiente los mismos programas genéticos promueven protección celular y mantenimiento (C. J. Kenyon, 2010).

La restricción calórica en la dieta, que permite estudiar mecanismos moleculares de detección de nutrientes y situación de estrés en una misma intervención, ha demostrado ser suficiente para extender la longevidad máxima en diversos organismos modelo. Por tanto, las rutas de señalización que regulen la respuesta ante dietas con retriicciones calóricas son perfectas candidatas para mostrar cómo puede alterarse el proceso de envejecimiento. Entre esas rutas de señalización encontramos las rutas insulina/IGF-1 (insulin-like growth factor-1), complejo TOR (target of rapamycin) y sirtuinas. Brevemente, la ruta señalizadora insulina/IGF-1 activa aguas abajo al factor de transcripción FOXO, el cual participa en la expresión de programas génicos para combatir el estrés, mantener la homeostasis de las proteínas y permite la extensión de la vida. En *C. elegans* mutaciones en genes homólogos de esa ruta dan lugar a gusanos que viven más del doble que la línea silvestre y en ratones heterocigotos IGF-1^{+/-} se aprecia un incremento del 33% en la longevidad en hembras y 16% en machos (C. Kenyon et al., 1993; Vitale et al., 2019). TOR forma parte de un mecanismo intracelular de integración de señales que interacciona con diferentes vías de señalización: factores de crecimiento (especialmente IGF), estrés, hipoxia, disponibilidad de aminoácidos y WNT. Su relación con la extensión de la vida está ampliamente documentada. Se han descrito alteraciones en la longevidad asociadas a multitud de mutaciones y cambios en la expresión de proteínas que componen el complejo TOR y proteínas cuyo papel en la transmisión de la señal se desempeña aguas arriba o aguas abajo de TOR (para ejemplos concretos véase la Tabla 1 de (Daniel S. Evans & , Wen-Chi Hsueh, 2011)). Finalmente, las sirtuinas han demostrado ser capaces de aumentar la longevidad en levaduras en un 70% mediante la sobreexpresión de SIR2 y en ratones SIRT6^{-/-} aparecen signos del envejecimiento a las 3 semanas del nacimiento (Grabowska et al., 2017). En el caso de las sirtuinas, aún se desconocen los mecanismos moleculares que las asocian a la longevidad.

Estos descubrimientos han desencadenado la búsqueda de terapias que permitan modular el proceso de envejecimiento, tratando de ralentizar la aparición de síntomas, disminuyendo la prevalencia de enfermedades asociadas en la población e incluso revertir o reparar los daños que presentan tejidos envejecidos. Es de especial relevancia el caso de la rapamicina, inhibidor de TOR capaz de aumentar levemente la longevidad en ratones cuando es aplicada en individuos adultos (Daniel S. Evans & , Wen-Chi Hsueh, 2011). Al ser un medicamento ya aprobado para uso en el tratamiento de algunos tumores, su comercialización para el tratamiento del envejecimiento en caso de que se demostrase su efectividad sería de relativa facilidad.

El uso de modelos animales ha sido crucial para estos hallazgos. Levaduras (*S. cerevisiae*), gusanos (*C. elegans*) y moscas (*D. melanogaster*) han sido ampliamente utilizados por su corta longevidad y la posibilidad de llevar a cabo experimentos de cribados masivos. Sin embargo, los modelos invertebrados no comparten todos los aspectos del envejecimiento humano y enfermedades asociadas, puesto que carecen de ciertos órganos y sistemas (sangre, huesos y sistema inmune adaptativo). Los

sistemas vertebrados, como por ejemplo ratón (*M. musculus*) y pez cebra (*D. rerio*), también se usan para el estudio del envejecimiento pero presentan grandes desventajas: una longevidad relativamente alta (longevidad máxima de 3-4 y 5 años, respectivamente) y unos costes de mantenimiento elevados (especialmente para ratones) (Harel et al., 2015).

Para salvar estos problemas, en el ámbito de la investigación del envejecimiento se comenzó a utilizar un nuevo modelo vertebrado, el pez killi (*Nothobranchius furzeri*). Esta especie fue encontrada en Zimbague y Mozambique. Viven en pequeños depósitos de agua transitorios formados durante la estación lluvios. Los peces killi adultos mueren tras el secado de la charca donde viven pero dejan embriones que se mantienen vivos enterrados en el barro durante la estación seca, permaneciendo en un estado quiescente (diapausa) hasta que la estación lluviosa llega otra vez. Este hábitat transitorio les ha hecho adaptarse a un crecimiento rápido y alcanzar una pronta madurez sexual (Cellerino et al., 2016). *N. furzeri* presenta varias ventajas en comparación con los modelos previamente descritos. En primer lugar, posee la menor esperanza de vida observada en vertebrados, alrededor de 13 semanas en la cepa menos longeva. Segundo, los peces pueden ser mantenidos en una densidad poblacional más alta que los ratones, disminuyendo el coste de mantenimiento y permitiendo el estudio de poblaciones. Además, los huevos resisten ambientes secos y pueden ser conservados a temperatura ambiente, en vista al mantenimiento de líneas en laboratorio. Por último, las hembras ponen una gran cantidad de huevos, por lo que las colonias crecen muy rápido (Mitchell et al., 2015). Su corto ciclo de vida dió lugar a multitud de experimentos para el estudio del envejecimiento de la especie, habiéndose constatado múltiples indicadores de que *N. furzeri* recopila muchos de los marcadores que se han demostrado relacionados con el proceso del envejecimiento. Añadiendo a eso una secuenciación del genoma de calidad, anotaciones genómicas y la existencia de protocolos para la transgénesis con CRISPR se puede apreciar la idoneidad de este modelo para el estudio del envejecimiento.

Durante las dos últimas décadas se ha producido un enorme desarrollo en las técnicas ómicas basadas en secuenciación masiva. El continuo abaratamiento y la estandarización de protocolos y metodologías ha propiciado su accesibilidad a muchos grupos de investigación. La ingente cantidad de datos que una o la combinación de múltiples técnicas ómicas aporta permite analizar las complejas relaciones existentes en los miles de genes y elementos reguladores que actúan en las células. El estudio del transcriptoma con la técnica RNA-seq permite el descubrimiento y cuantificación global de los transcritos en un solo experimento. El perfil transcriptómico de un tejido completo representa el estado celular subyacente del fenotipo estudiado en un instante concreto (Van den Berge et al., 2019). Por tanto, una serie temporal de perfiles transcriptómicos durante el envejecimiento puede arrojar luz en la biología de este proceso único. Para enriquecer los datos aportados por la técnica RNA-seq, es usual su combinación con análisis centrados en los elementos reguladores del genoma. Como bien es sabido, la expresión de los genes se encuentra meticulosamente controlada. En organismos eucariotas, uno de los factores que influye en dicha regulación es el grado de accesibilidad de la cromatina. Por efectos estéricos, las regiones de ADN accesibles a la maquinaria de transcripción

podrán ser expresadas, mientras que las regiones de ADN empaquetadas no lo serán. Diversas técnicas para el estudio de la regulación génica han sido desarrolladas. Estas incluyen experimentos para medir la unión de factores de transcripción o modificaciones en histonas (ChIP-seq, Chromatin Immuno-Precipitation sequencing) o ensayos para medir la accesibilidad de la cromatina (ATAC-seq; DNase-seq, DNase I hypersensitive sites sequencing; FAIRE-seq, Formaldehyde-Assisted Isolation of Regulatory Elements sequencing). Entre ellas, la técnica ATAC-seq produce una gran cantidad de información, haciéndola capaz de definir mapas de regiones de cromatina activa en diversidad de organismos modelo, estadios y condiciones (Yan et al., 2020). Ha ganado gran popularidad en los últimos años debido a que las regiones abiertas o activas de la cromatina pueden ser posibles elementos reguladores no codificantes, como promotores o potenciadores/silenciadores de la transcripción. Dadas las evidencias que sugieren la existencia de redes genéticas capaces de alterar el proceso de envejecimiento en los organismos, queremos investigar qué elementos del genoma alteran el proceso de envejecimiento y cómo se regulan. Para ello, se han realizado series temporales de experimentos de RNA-seq y ATAC-seq en el cerebro de pez killi, lo cual nos permitirá integrar la información sobre cómo se expresan los genes en el tejido e influencia de elementos reguladores en su actividad.

RESULTADOS Y DISCUSION

Con objeto de investigar las redes genéticas involucradas en el proceso de envejecimiento y enfermedades asociadas se analizaron las series temporales de datos de RNA-seq y ATAC-seq en las tres partes del cerebro (telencéfalo, mesencéfalo y romboencéfalo) de *N. furzeri* machos de la cepa MZM 04/10 a diferentes puntos temporales. Los puntos temporales se inician en la semana 8 tras la salida de los huevos de diapausa, momento en el que se estima que los individuos alcanzan la maduración sexual y se extiende hasta la semana 24, alcanzando las últimas semanas de vida de esta cepa (Terzibasi et al., 2008).

Análisis del transcriptoma

El control de calidad de las lecturas mapeadas en las muestras de RNA-seq se llevó a cabo sobre el conteo de lecturas por gen. Se empleó la representación de componentes principales (PCA) eliminando el 10% de los genes menos variables para observar la calidad de las réplicas biológicas (Figura suplementaria 1a). Aparecen patrones claramente visibles en telencéfalo y romboencéfalo en los que se aprecia que la transición entre puntos temporales tempranos y tardíos sigue una línea curva atendiendo a las dos componentes principales que más varianza recogen. Sólo las réplicas de la semana 16 se salen del patrón que siguen las demás, indicando que algún fenómeno biológico o artefactual ocurre en esa semana y que además es consistente en los tres tejidos. Para explorar aún más la adecuación de las muestras para los posteriores análisis, se estudió la correlación por parejas de muestras calculando el coeficiente de correlación de Spearman (Figura suplementaria 1b). Todas las muestras presentan alta correlación entre ellas no apareciendo ningún sesgo significativo. Por ello, consideramos que todas las muestras en las tres series temporales son adecuadas para su análisis poniendo especial atención en las muestras de la semana 16.

En primer lugar, se buscó extraer una visión general de a qué rutas metabólicas pertenecen los transcritos expresados durante el proceso de envejecimiento. Se estudió el enriquecimiento de grupos de genes curados y revisados que participan de las principales vías metabólicas con la herramienta GSEA (Subramanian et al., 2005) (Figura 1a). Los mecanismos relacionados con la división celular ('G2M CHECKPOINT', 'E2F TARGETS' y 'MITOTIC SPINDLE') correlacionan negativamente con la edad y de forma consistente en los tres tejidos, lo que implica una menor expresión de los genes relacionados con la mitosis. El análisis podría indicar que MYC puede tener relación con la disminución de la actividad en los programas proliferativos puesto que se conoce que MYC se encuentra estrechamente asociado a la proliferación celular y la apoptosis y se aprecia que las rutas diana de MYC ('MYC TARGETS V1' y 'MYC TARGETS V2') y uno de los coactivadores de MYC ('WNT BETA CATENIN SIGNALING') (Dang, 1999) correlacionan también negativamente con la edad. Además, aparecen varios de los marcadores de envejecimiento descritos en el cerebro: aumento de la expresión de programas para lidiar con las especies reactivas de oxígeno ('REACTIVE OXYGEN SPECIES PATHWAY'), respuesta inflamatoria ('INFLAMMATORY RESPONSE'),

alteración de la ruta señalizadora TOR ('MTORC1 SIGNALLING') y una disminución de las actividades de genes encargados en el control del plegamiento de proteínas y reparación del ADN ('UNFOLDED PROTEIN RESPONSE' y 'DNA REPAIR', respectivamente). A la hora de interpretar estos resultados se puede asumir que los grupos de genes no se superponen entre ellos (Figura 1b). En general, estos datos muestran que los tejidos más longevos están sometidos a estrés y que aparecen algunos de los marcadores del envejecimiento previamente descritos.

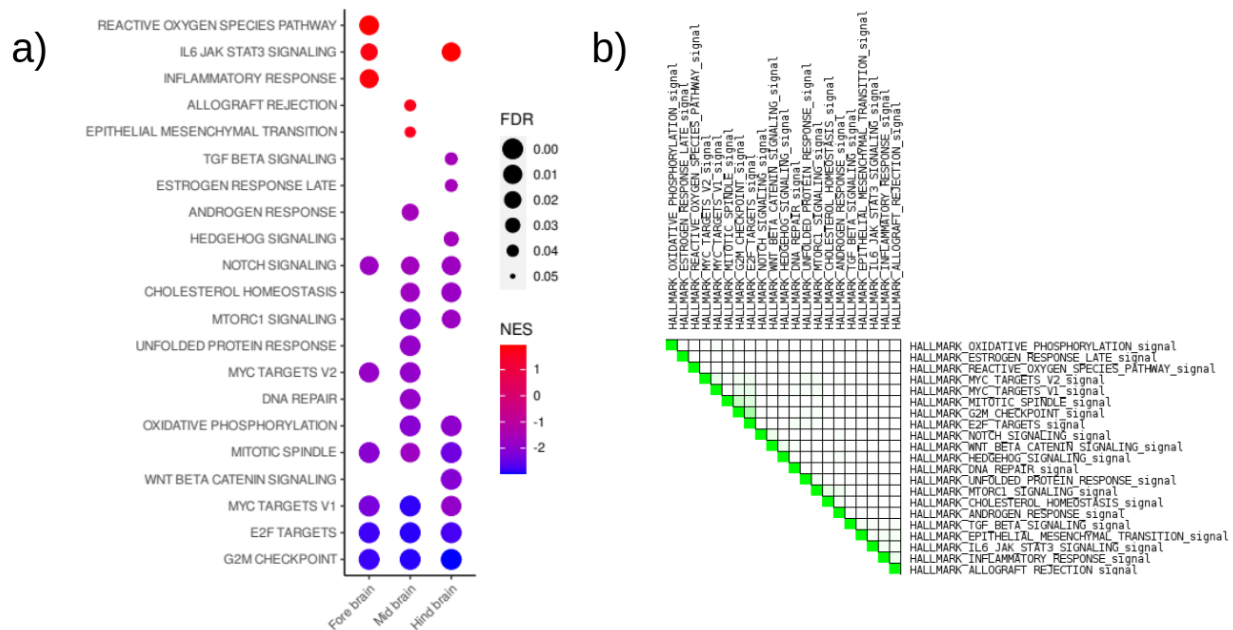


Figura 1. Enriquecimiento de las principales rutas metabólicas según GSEA. Fueron analizados los 50 grupos de genes pertenecientes a 'hallmark gene sets' de MsigDB en los tres tejidos. a) Se presentan los p-valores corregidos por BH de las rutas metabólicas significativas (q -valor $<0,05$) en al menos uno de los tejidos y están ordenados según el valor de enriquecimiento normalizado (NES). El tamaño del círculo representa el q -valor y su color el NES. $NES > 0$ implican una correlación positiva con la edad de la muestra y $NES < 0$ una correlación negativa. b) Heatmap representando el grado de solapamiento entre los genes de las rutas metabólicas con correlación significativa.

A continuación, se analizó el enriquecimiento de procesos biológicos definidos por términos GO en los genes diferencialmente expresados. Los contrastes para la búsqueda de significancia estadística se llevaron a cabo comparando cada punto de la serie temporal con el punto temporal más joven, la semana 8. De esta forma se pudo ver la evolución de los diferentes procesos biológicos a lo largo de la vida. En los tres tejidos puede apreciarse que se apagan rutas relacionadas con el desarrollo del sistema nervioso (Figura 2a), siendo un buen indicador de que se han conseguido seleccionar genes específicos de cerebro en puntos temporales alejados del periodo de desarrollo. También, se repite la aparición de programas de expresión relacionados con el ciclo celular regulados negativamente, corroborando los resultados aportados por GSEA. Específicamente, en mesencéfalo y romboencéfalo

disminuye la actividad de procesos relacionados con la organización intracelular y la adhesión célula-célula. Por otro lado, aparece un aumento de expresión conforme avanza el tiempo en genes relacionados con procesos de metabolismo biosintético como la síntesis de ribosomas, nucleótidos y péptidos (Figura 2b) en mesencéfalo y romboencéfalo, mientras que en telencéfalo no hay procesos biológicos significativos al comparar la semana 24 con la 8. Es interesante notar que algunos procesos biológicos son más significativos al comparar la semana 16 con la semana 8 que la propia semana 24, siendo especialmente visible en procesos cuya expresión aumenta en romboencéfalo. No parece existir ningún patrón entre las rutas que presentan esta característica, sin embargo no puede descartarse que exista un cambio programado en el transcriptoma del organismo.

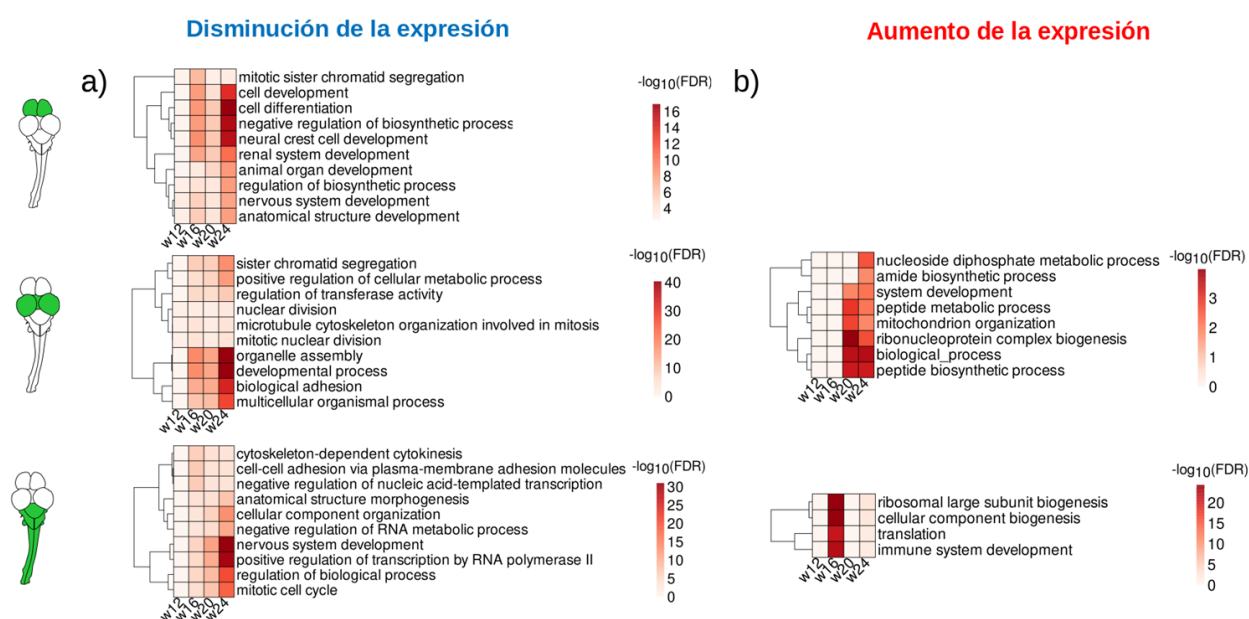


Figura 2. Cambios transcriptómicos asociados al envejecimiento en las diferentes partes del cerebro. Se tomó como referencia la semana 8. Se seleccionaron los 10 términos GO de la ontología procesos biológicos con menor q-valor al comparar la semana 24 con la semana 8. a) Procesos biológicos con $LFC < 0,75$. b) Procesos biológicos con $LFC > 0,75$. De arriba a abajo: telencéfalo, mesencéfalo y romboencéfalo.

Es posible que existan otras dinámicas de expresión más complejas, diferentes a un apagado o encendido progresivo en la expresión de grupos de genes. Para buscar esos perfiles, se aplicó el algoritmo c-means a los genes diferencialmente expresados (DEGs). Los DEGs fueron definidos por el test estadístico 'log likelihood ratio' implementado en DESeq2 (Love et al., 2014) (véase Métodos) de forma independiente para cada tejido. Se identificaron 2192, 2180 y 725 DEGs en telencéfalo, mesencéfalo y romboencéfalo (Figura 3a), respectivamente. Previamente a aplicar el algoritmo de clustering, se calculó el número de clusters óptimo siguiendo los métodos de número de clusters vacíos y distancia mínima entre centroides, resultando en un número de clusters óptimo de 9 para

cada tejido (Figura suplementaria 3). Los DEGs fueron agrupados asignando cada gen a un cluster. Tras la clusterización, se buscaron similitudes entre las dinámicas de clusters de diferentes tejidos, por lo que los centroides de los clusters fueron usados para calcular el coeficiente de correlación de Pearson entre cada pareja de clusters (Figura 3b), obteniendo así una medida no arbitraria que relacionase cada cluster con los demás. Aplicando k-means sobre la matriz de expresión se agruparon los centroides en 10 perfiles de expresión que se comparten en los diferentes tejidos (Figura 3c). En dos perfiles de expresión pueden verse comportamientos monotónicos: el perfil 1 y 8 desciende y asciende gradualmente, respectivamente. El resto de clusters muestran patrones más complejos como un pico mínimo en forma de V en el perfil 5 o su patrón complementario con un pico máximo en la semana 16 en el perfil 6. Finalmente, se investigó si los diferentes perfiles de expresión correspondían a procesos biológicos diferentes y si podían observarse diferencias entre los diferentes tejidos del cerebro. Para ello se analizaron los términos GO significativos (q -valor $<0,05$) asociados a los genes de cada cluster (Figura 3d). En general, las rutas metabólicas detectadas en el análisis para cada perfil de expresión se conservan en el resto de tejidos, observándose múltiples triadas de puntos en las mismas posiciones que representan que aparecen como significativos los mismos términos GO. Además, aunque previamente se ha mostrado que genes relacionados con el desarrollo aparecen enriquecidos en dinámicas que siguen patrones de apagado, aquí podemos ver que casi todos los perfiles de expresión presentan términos GO relacionados con el desarrollo y la diferenciación. Esto podría indicar que las rutas involucradas en el desarrollo podrían tener otro papel durante el proceso de envejecimiento. Por otro lado, encontramos procesos relacionados con las células neuronales en los perfiles 3, 5 y 9. Dado que el perfil 3 sigue una dinámica de decaimiento temprano, teniendo un pico en la semana 8, es posible que el cluster contenga genes involucrados en el desarrollo neuronal; haciendo especialmente interesante los perfiles 5 y 9, los cuales presentan una leve correlación positiva al compartir un pico en forma de valle en la semana 16. A ello habría que sumar que, tanto el perfil 5 como el 9, agrupan los términos relacionados con sinapsis. Finalmente, cabe destacar la presencia aislada a un tipo de dinámica, un descenso gradual de la expresión, de los genes relacionados con la división celular y el metabolismo de ADN.



Figura 3. Análisis de las dinámicas de expresión. a) Heatmaps de los DEGs identificados en cada uno de los tejidos (de izquierda a derecha: telencéfalo, mesencéfalo y romboencéfalo). Se presentan los valores estandarizados a partir de la normalización de DESeq2 b) Matriz de correlación entre los 27 centroides de los clusters obtenidos. Aplicando k-means podemos agrupar clusters de diferentes tejidos en perfiles de expresión similares. c) Dinámicas seguidas por cada uno de los perfiles de expresión. Se representan los centroides de cada cluster tras estandarizar (véase el apartado Dinámicas de expresión, Métodos). d) Términos GO asociados a cada perfil de expresión. Los términos GO fueron agrupados en categorías usando expresiones regulares. Cada categoría GO engloba tantos términos GO como se especifica al lado de la abreviatura. Las columnas agrupan los clusters según el perfil de expresión al que pertenecen y, dentro de ellos, el tejido al que pertenecen. La escala de color representa el q-valor de cada término GO. Las abreviaturas son (de arriba a abajo): Cell division, Development, Differentiation, DNA, Ion transport, Neuron, Ribosome, RNA y Synapse.

Análisis de elementos reguladores

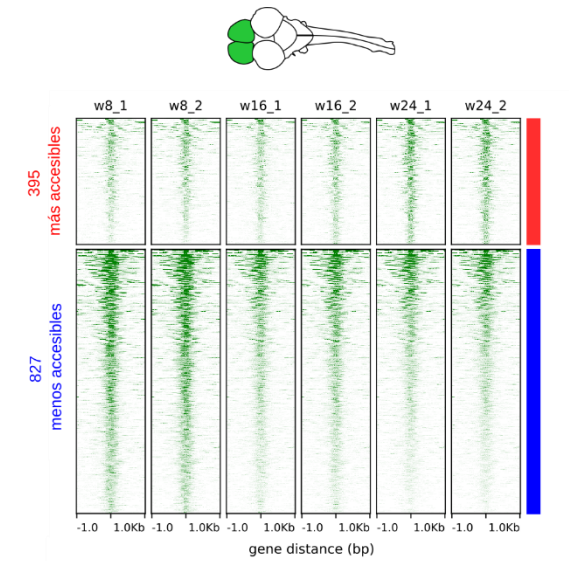
Tras el mapeo e identificación de picos ('peak calling') de las lecturas de ATAC-seq se identificaron las regiones abiertas de cromatina (RACs), entre las cuales se encuentran promotores y elementos reguladores de la transcripción. Las RACs que mostraron alta consistencia entre réplicas biológicas según el método IDR (Q. Li et al., 2011) fueron utilizadas en los siguientes análisis. Se unieron las coordenadas de las RACs presentes en todas las muestras pertenecientes a cada tejido para formar un consorcio de RACs, lo que permitió contar el número de lecturas por pico en cada punto temporal. Dada la sensibilidad de la técnica ATAC-seq, la cantidad de RACs tras el preprocesado es enorme, alrededor de 100.000 RACs en telencéfalo y mesencéfalo y 90.000 RACs en romboencéfalo. Por tanto, se optó por una estrategia de evaluación de la calidad de las muestras distinta a la seguida en RNA-seq. En primer lugar, las RACs diferencialmente accesibles (dRACs) fueron calculadas usando el test 'quasi-likelihood F-test' implementado en EdgeR (Robinson et al., 2009) en lugar del test LRT usado anteriormente, ya que permite un control del ratio de error más robusto cuando el número de réplicas biológicas es pequeño, según la documentación del paquete EdgeR. Aún así, la cantidad de dRACs obtenidas era pequeña y los posteriores análisis de esas listas pobres (no se muestra). Por tanto, se buscó aumentar la cantidad de dRACs por los análisis diferenciales. Las dRACs estadísticamente significativas se definen como aquellas que presentan un p-valor corregido según Benjamini Hochberg (BH) $<0,05$. La corrección de BH es muy útil para controlar la proporción de falsos positivos, sin embargo, lo hace a costa de desechar una gran proporción de verdaderos positivos. A mayor cantidad de elementos testados por BH, mayor será la cantidad de verdaderos positivos eliminados. Por tanto, filtrar los datos de ATAC-seq antes de aplicar la corrección de BH reducirá la cantidad de testeos aumentando el número de dRACs consideradas significativas. En general, las dRACs con bajo número de lecturas presentarán menos evidencias a la hora de estimar si existen diferencias significativas, motivo por el cual es común aplicar filtros por bajos conteos. Aquí, se combinaron los métodos de filtros por bajos conteos presentes en DESeq2 y EdgeR para encontrar

el filtro óptimo que maximice el número de dRACs significativas (Figura suplementaria 4). A modo de ejemplo, en mesencéfalo el máximo número de dRACs se obtiene al filtrar todos aquellos RACs con unos conteos inferiores al cuantil 0,7 de los datos, lo que corresponde a un límite en conteos por millón (CPM) de 9,58. Si utilizáramos el filtro por defecto de 2 CPM, el número de dRACs sería aproximadamente la mitad. Con esta estrategia se obtuvieron 1222, 849 y 720 dRACs en telencéfalo, mesencéfalo y romboencéfalo respectivamente (Figura 5a). Los dRACs fueron utilizados para evaluar la calidad de las muestras con la representación PCA, el cálculo de los coeficientes de correlación de Spearman por parejas (Figura suplementaria 5) y en análisis posteriores. Los puntos temporales de las diferentes semanas consiguen separarse por réplicas en la PCA excepto una de las réplicas de la semana 16 en romboencéfalo. Dicha muestra presenta una correlación levemente superior con muestras de la semana 24 que con su réplica biológica, sin embargo no fue considerado un valor tan extremo como para no tenerla en cuenta, pudiendo achacarse a la variabilidad inter-individuos durante el proceso de envejecimiento.

Tras la identificación de regiones cromosómicas que parecen estar sujetas a regulación según cambios significativos en su apertura, se investigó la presencia de motivos de unión al ADN en esas regiones. Para ello se utilizó el conjunto de herramientas y base de datos de motivos de Homer. Para cada tejido, las dRACs fueron divididas en dos grupos: dRACs cuya accesibilidad aumenta con la edad de la muestra y dRACs cuya accesibilidad disminuye. En ellos se analizó la presencia de motivos de unión de factores de transcripción con evidencia experimental. Sólomente en el grupo de dRACs cuya accesibilidad aumenta en romboencéfalo no se pudo encontrar motivos significativos, probablemente debido al bajo número de dRACs para estudiar. En telencéfalo y mesencéfalo aparecen motivos de unión similares que siguen una dinámica de apertura similar, a los que se unen los factores de transcripción KLF14, MAZ y CTCF. KLF14 es un co-represor transcripcional del receptor TGF-beta II cuando es activado por TGF-beta, miembro de una de las principales vías de señalización de diferenciación, proliferación y homeostasis de múltiples tejidos (Truty et al., 2009; Weiss & Attisano, 2013). MAZ también está relacionado con la proliferación celular y apoptosis, pues es un dedo de zinc asociado a MYC, además de actuar como factor de transcripción independiente bajo el nombre de Pur-1 que activa los promotores de insulina. Por su parte, CTCF es un dedo de zinc que participa del plegamiento 3D del genoma para facilitar la interacción de elementos cis reguladores distantes, mediante la formación de dominios topológicamente asociados (TADs) favoreciendo la adecuada ejecución de programas genéticos durante el desarrollo embrionario (Franke et al., 2020). Sorprendentemente, hace poco se ha relacionado a MAZ con la organización tridimensional del genoma y se ha relacionado con la interacciones que ocurren dentro de los TADs y el arresto de moléculas de cohesina (Xiao et al., 2021), proteínas presentes en las regiones que delimitan los TADs. En telencéfalo se detecta además la pérdida de accesibilidad del motivo de unión de EN1 y la apertura del motivo de BORIS, un parálogo de CTCF en mamíferos (Pugacheva et al., 2015), y OLIG2. Por otro lado, en mesencéfalo disminuye la accesibilidad del factor NFY y se permite la unión al ADN de RUNX y ABF1. En romboencéfalo, aparecen como significativos motivos de unión muy similares que siguen una dinámica de pérdida de accesibilidad. Los tres factores de transcripción que

interaccionan con dichos motivos, OLIG2, NEUROG2 y ATOH1, están involucrados en procesos de diferenciación de células neuronales. Es interesante notar que el motivo para OLIG2 presenta dinámicas diferentes en telencéfalo y en romboencéfalo, pudiendo indicar la actuación de los mismos elementos reguladores en programas genéticos diferentes.

a)

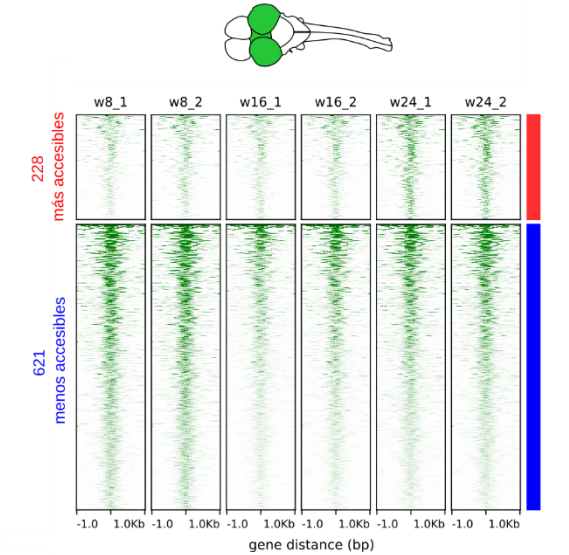


Factor $-\log_{10}(\text{p-valor})$ Motivo de unión

CTCF	16	ATAGTCCCACTCTGGGSA
BORIS	13	GGGCGCCCTCTGGG
OLIG2	9	ACCATCTGTT

KLF14	30	CTGGGCGGSC
MAZ	29	GGGCGGG
EN1	29	GGTAATTAS

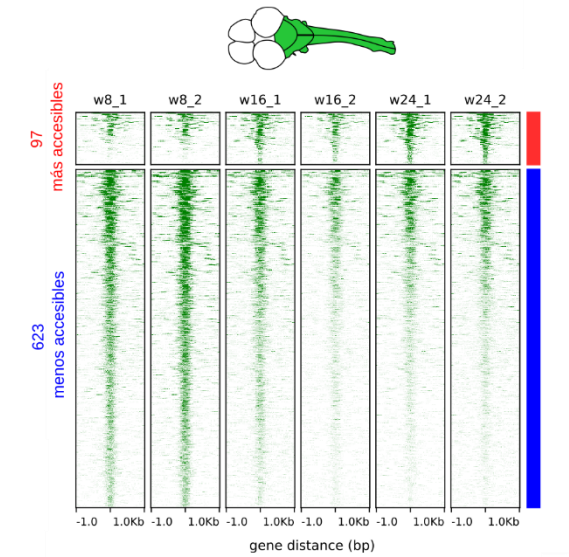
b)



CTCF	7	ATAGTCCCACTCTGGGSA
RUNX	7	AAACCAAGG
ABF1	6	CACGTGC

KLF14	34	CTGGGCGGSC
MAZ	33	GGGCGGG
NFY	33	ACCAATGAG

c)



OLIG2	34	ACCATCTGTT
NEUROG2	31	ACCATCTGTT
ATOH1	30	GTACACCTGCT

Figura 4. Motivos de unión al ADN en los dRACs según Homer. En (a) telencéfalo, (b) mesencéfalo y (c) romboencéfalo. En los heatmaps se ordenan las muestras en semana 8, 16 y 24 y se representa la accesibilidad a ± 1 kb de distancia del centro de cada pico. Los dRACs se dividieron en regiones cuya accesibilidad aumenta o disminuye durante el envejecimiento. A la derecha se muestran los 3 motivos de unión con menor p-valor, su p-valor y el factor de transcripción que se une a dicho motivo.

Es notable que en los motivos descubiertos, especialmente en aquellos cuyos dRACs incrementan su accesibilidad, el grado de evidencia para evaluar la presencia de un motivo es bajo. Está influido por la pequeña cantidad de dRACs por grupo. Para compensarlo, se decidió validar los resultados paralelamente utilizando otro método para la búsqueda de motivos de unión. Se utilizó la herramienta Gimme motifs (Bruse & Heeringen, 2018) sobre los grupos de dRACs, la cual escanea motivos diferenciales y combina los resultados de varios métodos de regresión en un z-score final que representa la presencia de un motivo en una muestra. Puesto que dicha herramienta no reporta p-valores, otras estrategias de priorización fueron utilizadas. Se recurrió a diversos análisis de asociación. Por un lado, se descargaron y filtraron tres bases de datos relacionadas con el envejecimiento (Tacutu et al., 2018): GenAge Human (genes relacionados con envejecimiento en humanos), CellAge (genes asociados a senescencia) y DrugAge (medicamentos que alteran la longevidad en organismos modelo). Estas bases de datos fueron utilizadas para calcular la asociación entre las listas de factores de transcripción aportadas por Gimme motifs y diferentes aspectos del envejecimiento. Por otro lado, se estudió si los factores de transcripción con motivos de unión diferenciales podían explicar los cambios transcriptómicos descritos anteriormente. Para ello, se utilizó la herramienta TFEA.ChIP (Puente-Santamaria et al., 2019). Todo ello se ilustra en la Figura 5. Entre los factores de transcripción asociados a motivos diferenciales en telencéfalo según Gimme motifs, encontramos tres factores asociados a envejecimiento (EGR1, ARNTL y NR3C1, p-valor test de Fisher = 0,0014) y dos asociados a senescencia (EWSR1 y NANOG, p-valor test de Fisher = 0,018). De ellos, EGR1 y NANOG, factores involucrados en proliferación y diferenciación celular, siguen una dinámica de apagado, lo cual correlaciona con los anteriores resultados al indicar una disminución en la capacidad proliferativa. En cambio, en dinámicas de apertura aparece ARNTL, factor cuya depleción en ratones provoca una reducida esperanza de vida y síntomas de envejecimiento prematuro (Kondratov et al., 2006), además de incrementar los niveles de especies reactivas de oxígeno en telencéfalo (Ali et al., 2020). En mesencéfalo aparecen una gran cantidad de factores relacionados con envejecimiento (EGR1, RELA, ATF2, STAT5A y BRCA1, p-valor test de Fisher = $1,02 \times 10^{-6}$), senescencia (GATA4 y BRCA1, p-valor test de Fisher = 0,013), diana de medicamentos que alteran la longevidad (GATA4, RELA y BRCA1, p-valor test de Fisher = 0,02) y asociados a los DEGs (ELK1, STAT5A, ZBTB33, ATF2, EGR1 y RELA, q-valores test de Fisher $< 0,05$). Entre los factores enriquecidos positivamente en etapas más jóvenes aparecen nuevamente EGR1 y ELK1, compartiendo la misma dinámica que en telencéfalo, sin embargo, en mesencéfalo aparece evidencia transcriptómica que verifica la actividad de ambos factores. Además, encontramos enriquecido en la semana 8 a RELA, una de las subunidades de NF- κ B. Existen evidencias de su

función en neuronas y células gliales, donde la sobreactivación de RELA promueve apoptosis tras daños isquémicos (Lanzillotta et al., 2015). Durante la semana 16 hay un grupo de motivos que aparecen enriquecidos de forma transitoria, a los que se unen los factores STAT5A y ATF2. El papel de ambos ha sido estudiado en el cerebro, participando ATF2 en la regulación de la supervivencia neuronal e inflamación (Liu et al., 2020) y la ruta de señalización de STAT5A en la ingesta de alimento, grasa corporal y homeostasis de glucosa (Furigo et al., 2020). Destaca el motivo de unión de los factores ZBTB33 y BRCA1 pues posee todas las anotaciones aquí estudiadas. Nótese que BRCA1 no tiene dominios de unión a ADN, lo que sugiere la interacción de BRCA1 con el dedo de zinc ZBTB33 para realizar su función como factor de transcripción (Wang et al., 2012), aunque no hay evidencias de las redes transcripcionales que ambos controlarían. No obstante, sí que se ha asociado a ZBTB33 (o su sinónimo KAISO) con el arresto del ciclo celular y apoptosis mediado por P53 (Choi et al., 2019) y la depleción de BRCA1 en cerebro con el aumento de roturas de ADN sin reparar, suponiendo multitud de problemas cognitivos a los organismos KO (Suberbielle et al., 2015). Finalmente, en romboencéfalo se observa enriquecimiento para genes asociados a senescencia (NANOG y IRF3, p-valor test de Fisher = 0,003). Al igual que en telencéfalo, el motivo NANOG disminuye su accesibilidad con la edad. Por su parte, el motivo de IRF3, factor asociado a la respuesta inflamatoria en cerebro (Canivet et al., 2018), aumenta su accesibilidad siendo posible su relación con los fenómenos inflamatorios que ocurren a edades tardías.

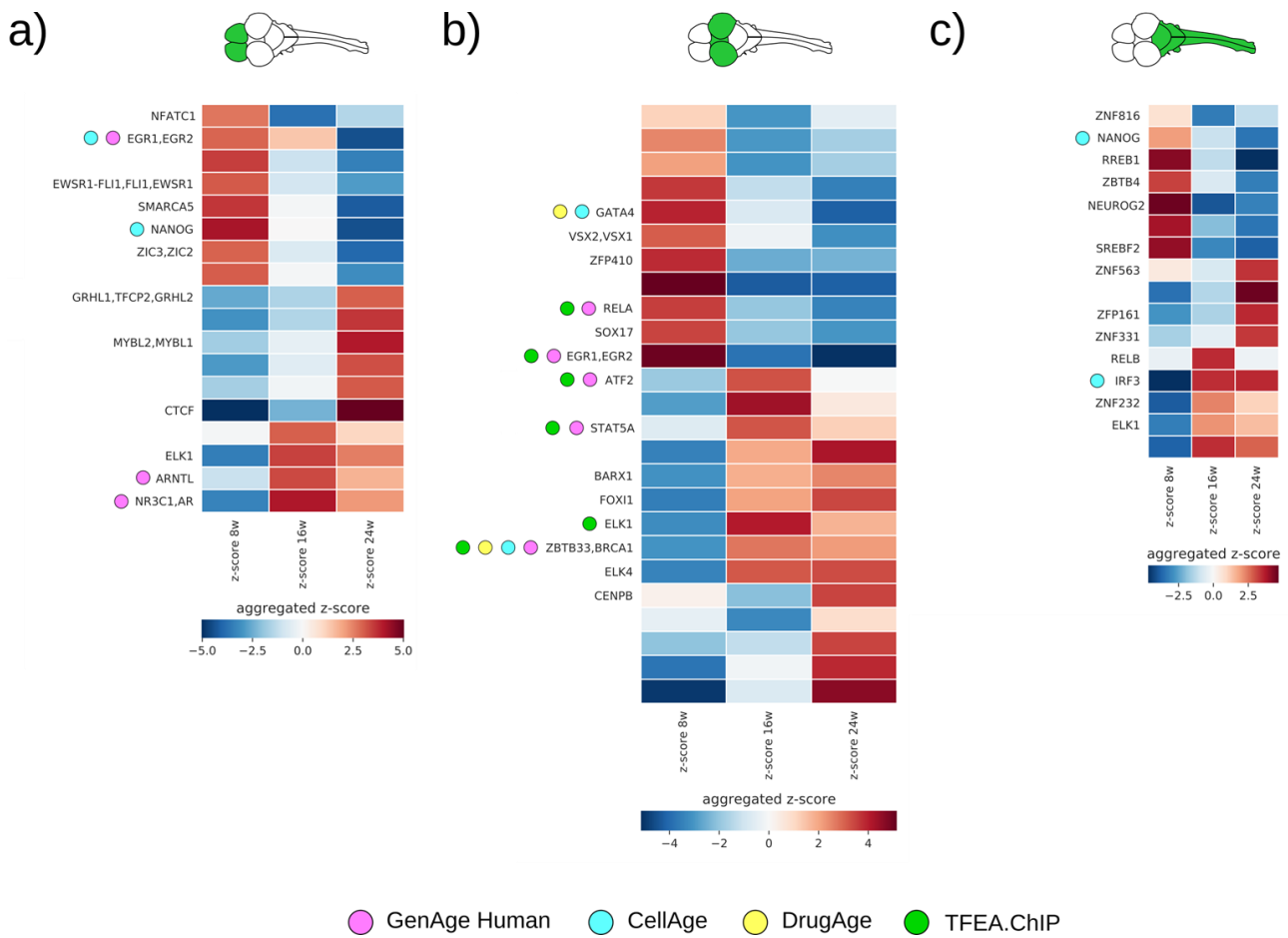


Figura 5. Factores de transcripción que se unen a motivos de unión diferenciales según Gimme motifs en (a) telencéfalo, (b) mesencéfalo y (c) romboencéfalo. Los z-score representan la señal en cada una de las muestras. Se muestran factores de los que se posee evidencia experimental de unirse a los motivos de ADN. Se añaden anotaciones sobre genes relacionados con envejecimiento (GenAge Human), genes asociados a senescencia (CellAge), genes que son diana de medicamentos que alteran la longevidad en organismos modelo (DrugAge) y factores de transcripción enriquecidos dados los DEGs (TFEA.ChIP).

Integración de los datos de RNA y ATAC-seq

La asociación de las dRACs a los sitios de inicio de la transcripción de los genes más cercanos fue utilizada para extraer listas de genes potencialmente regulados por esas regiones de cromatina activa. Se buscaron genes en común entre dichas listas y los DEGs definidos según el test LRT para cada tejido. Estos genes pueden ser afectados por elementos reguladores importantes en el proceso de envejecimiento, pues muestran a la vez dinámicas de expresión alteradas y regiones de cromatina cuya accesibilidad cambia en su vecindad (promotores o enhancers). Han sido detectados 202, 159 y 47 genes que cumplen estas condiciones en telencéfalo, mesencéfalo y romboencéfalo, respectivamente (Figura 6a). Análisis de enriquecimiento en términos GO revela la presencia de genes involucrados en el desarrollo en todos los tejidos y términos relacionados con la regulación de procesos biológicos ('regulation of cellular metabolic process' en telencéfalo y 'biological regulation' en mesencéfalo), lo que demuestra que en efecto se han seleccionado genes que controlan otros programas genéticos. De hecho, en romboencéfalo puede verse que la generación y diferenciación de neuronas se ve comprometida (Figura 6b).

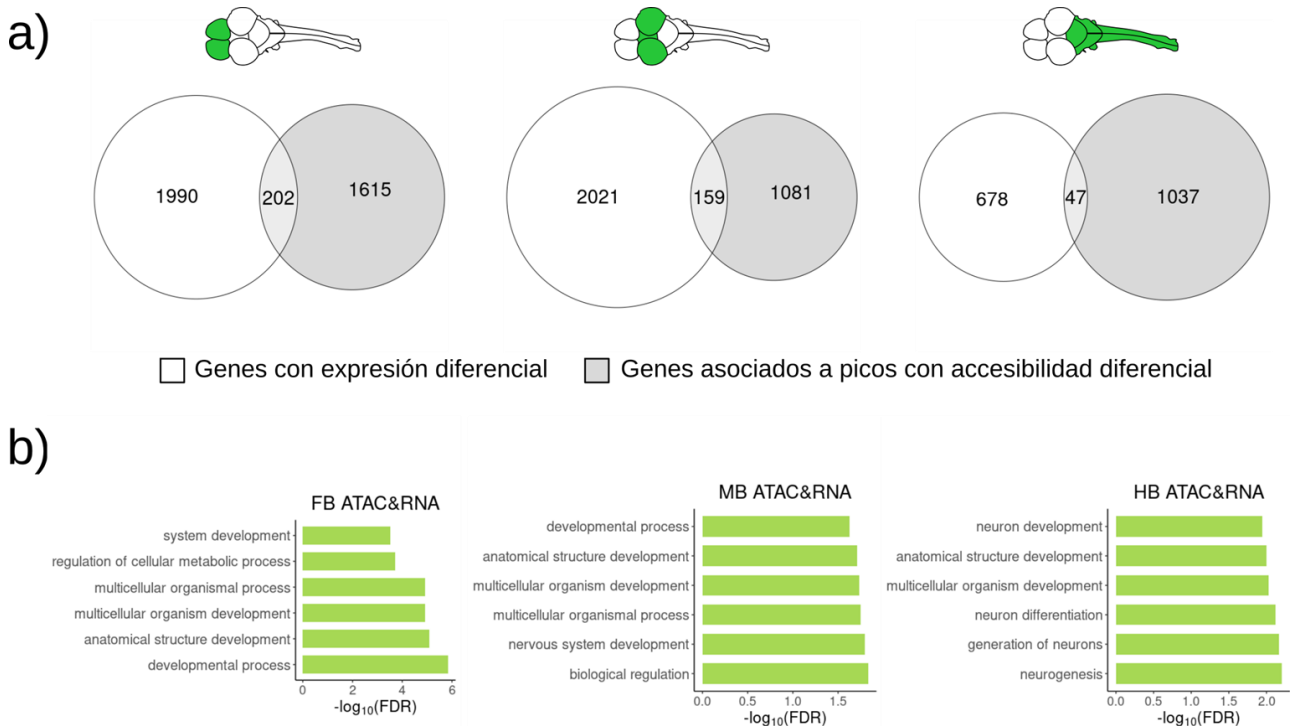


Figura 6. Integración de los genes con expresión diferencial y dRACs asociadas. a) Diagramas de Venn. b) Enriquecimiento en términos GO para los genes que interseccionan. Se muestran los 5 términos con menor q-valor. De izquierda a derecha: telencéfalo, mesencéfalo y romboencéfalo.

Por otro lado, se quiso estudiar la influencia de la accesibilidad de los promotores en la expresión génica en la semana 24. Se definió la región promotora como las 500 bp que rodean el sitio de inicio

de la transcripción (TSS) de cada gen, relacionando así las lecturas de ATAC-seq con las lecturas de RNA-seq para cada gen. La variabilidad observada en la accesibilidad de los promotores no puede explicar la variabilidad de la expresión génica ($R^2 = 0,14$, $R^2 = 0,14$ y $R^2 = 0,12$ en telencéfalo, mesencéfalo y romboencéfalo, respectivamente) (Figura 7a y Figura suplementaria 6a,c), lo cual es esperable pues un aumento en la accesibilidad de los promotores no necesariamente significaría aumento de la actividad del gen. También pueden estar asociado con represión de los genes. Para analizar algunos de los aspectos de la relación entre ATAC-seq y RNA-seq se dividieron los genes en grupos en función de su expresión y accesibilidad de los promotores (Figura 7b y Figura suplementaria 6b,d) y se aplicaron análisis de enriquecimiento funcional. Los grupos de genes más numerosos fueron los genes con alta accesibilidad y expresión (HA-HE) y los genes con baja accesibilidad y expresión (LA-LE). En el grupo HA-HE (Figura 7c), presumiblemente aquellos genes que tienen promotores que activan su expresión, encontramos genes relacionados con procesos metabólicos en los tres tejidos. También se observa que se llevan a cabo procesos de desarrollo, característica especialmente común en telencéfalo. No obstante, es en el grupo LA-LE (Figura 7e) donde los principales términos GO encontrados están asociados a procesos de desarrollo y organogénesis de forma consistente en todos los tejidos, siendo esperado este resultado pues se entiende que en las últimas etapas de la vida no se expresan ni activan promotores asociados a esta función. En el grupo de genes con alta accesibilidad y baja expresión (HA-LE) (Figura 7d) se encuentran procesos relacionados con el ciclo celular, indicando la posibilidad de que factores represores interactúen con los promotores de estos genes. Finalmente, en el grupo de genes con baja accesibilidad y alta expresión (Figura 7f) sólo encontramos enriquecimiento de rutas biológicas en telencéfalo, estando relacionadas con la secreción y movilidad de sustratos en el interior de la célula.

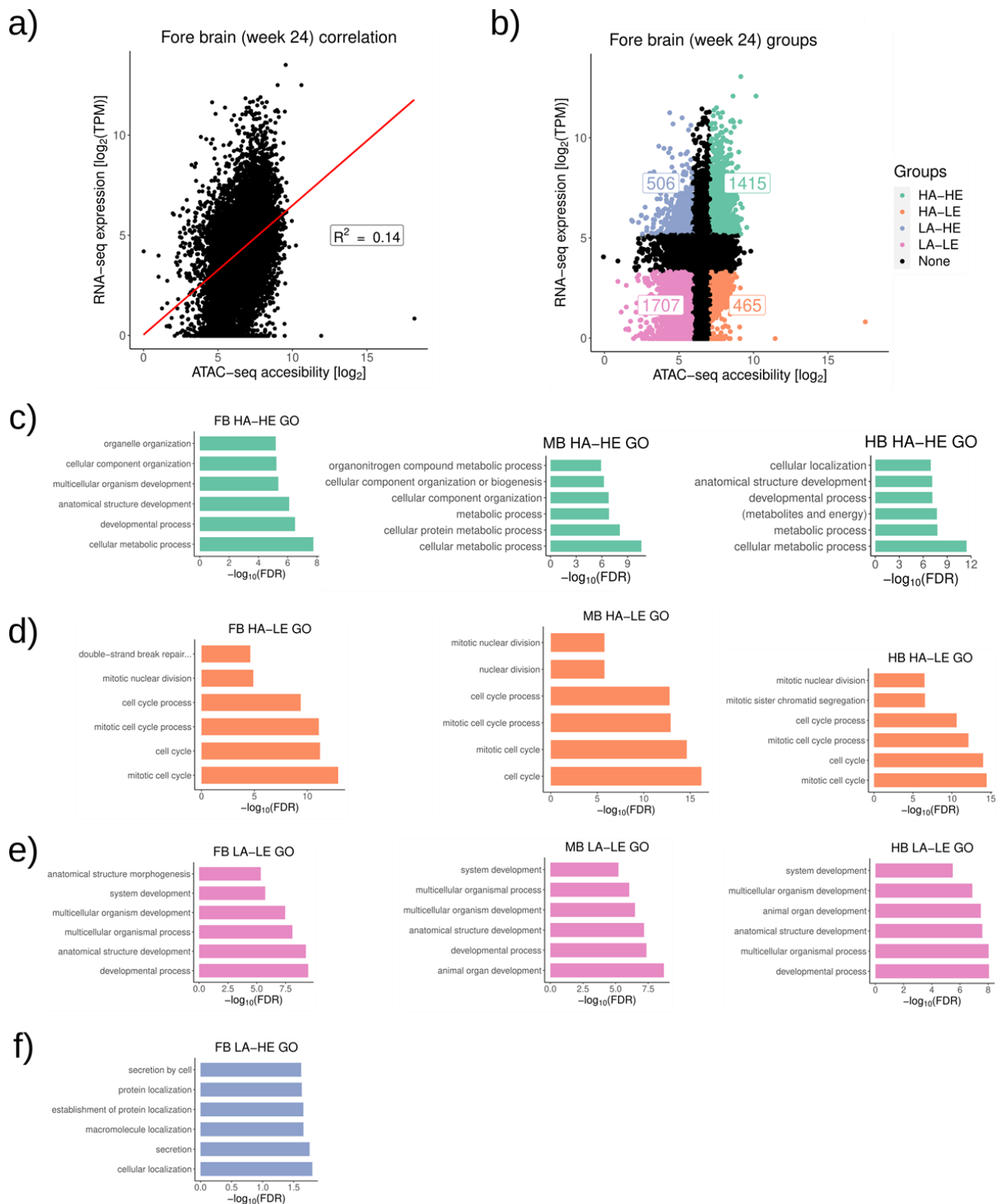


Figura 7. Accesibilidad de los promotores en la semana 24. a) Diagrama de puntos mostrando que la variabilidad observada en la accesibilidad de los promotores no explica la variabilidad en la

expresión de los genes. Se muestra el coeficiente de determinación (R^2). b) Los genes se agrupan basándose en el nivel de accesibilidad del promotor y expresión. En (a) y (b) se muestra telencéfalo como tejido representativo (véase Figura suplementaria 6 para el resto de tejidos). c), d), e) y f) Enriquecimiento de procesos biológicos para los genes de los grupos HA-HE, HA-LE, LA-LE y LA-HE, respectivamente. Se muestran los 5 términos con menor q-valor. De izquierda a derecha: telencéfalo, mesencéfalo y romboencéfalo.

Genes con dinámicas específicas de tejido

Hasta ahora se ha podido observar que algunos procesos biológicos están conservados durante el envejecimiento en las tres partes del cerebro. Sin embargo, dado el proceso de especialización regional que sufre este tejido, estas partes del tejido poseen diferencias anatómicas, de plasticidad y actividad neuronal y de las funciones que en otras partes del organismo controlan (Liscovitch & Chechik, 2013). Por ello, es posible que el proceso de envejecimiento afecte de forma diferente a cada una de ellas. Con los datos de RNA-seq se han buscado genes que presenten una dinámica tejido-específica en los puntos temporales observados (véase el apartado Identificación de genes específicos de tejido, Métodos). Se identificaron 121 genes que siguen patrones lineales diferentes en al menos uno de los tejidos, de los cuales 36 genes presentan al menos un pico de ATAC-seq asociado con accesibilidad diferencial. Algunos de ellos han sido estudiados en ensayos KO en diferentes organismos modelo demostrándose su relación con una correcta fisiología de neuronas y células accesorias en el cerebro (SLIT1B (Conceição et al., 2019), PAX2A (Conceição et al., 2019), NAP1L1 (Lee et al., 2017), KCNN1A (Autuori et al., 2019), entre otros), lo cual indica que podrían encontrarse en esta lista genes previamente no asociados con el deterioro de funciones fisiológicas. Cabe destacar la aparición en esta lista de dos genes que codifican para cadherinas, CDH6 y CDH13, y cuyos cambios de expresión son similares (Figura 8a,b). Ambos disminuyen su expresión en romboencéfalo mientras que aumentan su expresión en telencéfalo y mesencéfalo a lo largo de la vida. Además, CDH6 tiene asociado un pico de ATAC-seq (Figura 8c) diferencial en romboencéfalo que pierde accesibilidad a medida que envejecen las muestras, pudiendo ser un elemento regulador activador específico de este tejido. CDH6 es una de las proteínas de adhesión extracelular que se localizan en el espacio presináptico. En ratón, neuronas KO para CDH6/9/10^{-/-} no son capaces de establecer sinapsis en células HEK, mientras que ratones con dichas mutaciones en heterocigosis sí (Yamagata et al., 2018).

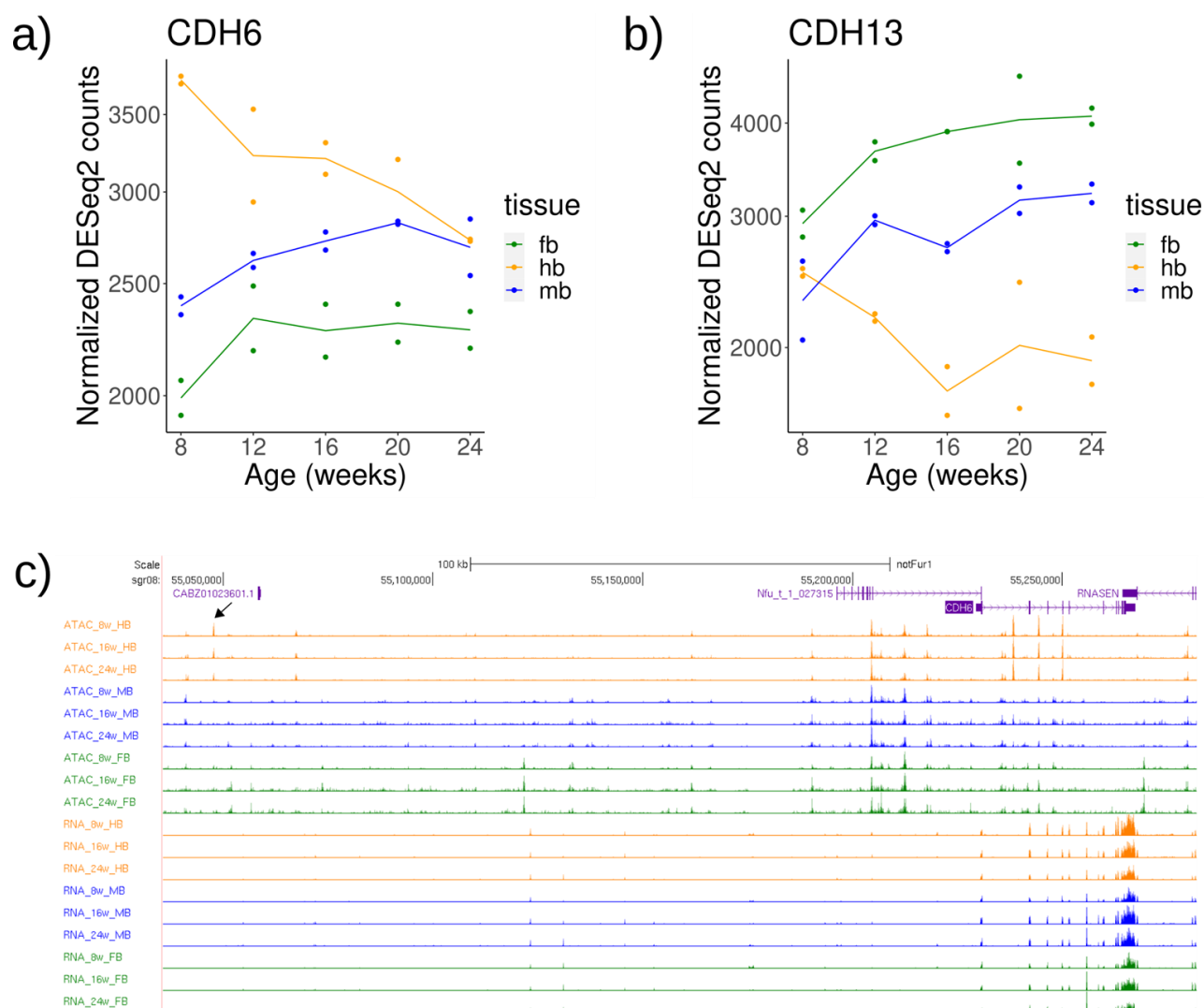


Figura 8. Ejemplo de genes con dinámicas de expresión tejido-específico. Dinámicas de expresión de (a) CDH6 y (b) CDH13. Se usa una escala de conteos logarítmica. c) Visualización en UCSC del gen CDH6 y el pico de ATAC-seq diferencial (señalado con una flecha).

Genes de desarrollo en envejecimiento

En los experimentos llevados a cabo ha sido consistente la observación de que los programas genéticos involucrados en desarrollo sufren cambios a lo largo de la vida de *N. furzeri*. Como ya se ha señalado, es de esperar que procesos relacionados con la organogénesis o la maduración aparezcan reprimidos, pues las muestras biológicas comenzaron a recogerse tras la maduración sexual. De hecho, es muy probable que este tipo de procesos hayan sido recogidos en nuestros análisis. Por ejemplo, el cluster 3 de los perfiles de expresión (Figura 3c,d) sufre un decaimiento temprano en la semana 8 seguido de una estabilización en el resto de la serie temporal. También, al estudiar la accesibilidad de los promotores, se ve que los genes que tienen poca expresión y cuyos promotores

son poco accesibles están relacionados con organogénesis y procesos anatómicos (Figura 7e). Ambos apuntan a que los programas de desarrollo están reprimidos. Sin embargo, existen otras evidencias que indican que hay otros genes relacionados con el desarrollo que no se reprimen, incluso aumentan su expresión en las últimas etapas de la vida. Es el caso del perfil de expresión 9, que presenta un pico de expresión positiva en las semanas 20 y 24 (Figura 3c,d). Es más, llama la atención que casi todos los perfiles de expresión definidos contienen genes involucrados en el desarrollo. Esto no puede ser artefactual, ya que previamente a realizar el clustering los genes mostraron tener dinámicas de expresión significativas en toda la serie temporal. Por tanto, se cree que algunos de los programas genéticos relacionados con el desarrollo podrían tener efectos diferentes durante el proceso de envejecimiento. Esta teoría sobre el envejecimiento se engloba y extiende bajo el nombre de pleiotropía antagonista.

Siguiendo esta línea, hay estudios que sugieren que algunos cambios transcriptómicos ocurridos en el envejecimiento se originan en el desarrollo. Somel et al. (Somel et al., 2010) midieron ARNm y ARNm_i (ARNm de interferencia) en el cerebro de humanos y macacos durante su vida. Descubrieron que la mayoría de genes y ARNm_i diferencialmente expresados durante el envejecimiento también estaban diferencialmente expresados durante el desarrollo. En ratones, Lui et al. (Lui et al., 2010) también estudiaron los perfiles transcriptómicos durante el envejecimiento en riñones, hígado y pulmón. Encontraron que muchos cambios en la expresión de los genes asociados a la edad se originaron en el desarrollo, particularmente en el período juvenil en el que el crecimiento se ralentiza. Ello supone la existencia de programas genéticos coordinando crecimiento y desarrollo que persisten en envejecimiento.

Concretamente, en relación al crecimiento, nuestros análisis demuestran que los genes relacionados con el ciclo celular disminuyen su expresión de forma constante desde la semana 12 (cluster 1 en la Figura 3c,d). Adicionalmente, aparece una disminución de la accesibilidad de motivos de unión al ADN de múltiples factores de transcripción que promueven la proliferación: KLF14, MAZ, EGR1 o NANOG (Figuras 4 y 5). Puesto que para el crecimiento de los tejidos se requieren nuevas células, la bajada de la tasa proliferativa demuestra que las series temporales estudiadas en este trabajo abarcan el período de no crecimiento de los organismos. Esto apoya la teoría de la pleiotropía antagonista, pues ¿cómo se concilia la idea de presencia de actividad en genes de desarrollo cuando la capacidad proliferativa está mermada? De nuevo, debe ser que los genes de desarrollo participan de otras funciones durante el envejecimiento. En conjunto, estos resultados sugieren que algunos aspectos del envejecimiento se encuentran codificados en el genoma formando parte de los procesos de desarrollo.

Ausencia del factor sexo en los experimentos

Nótese que lo presentado en este trabajo no tiene en cuenta el factor sexo debido a que todas las muestras provienen de machos. El estudio fue diseñado así para controlar que las diferencias observadas fueran debidas exclusivamente al envejecimiento. Además, para incluir hembras en los

análisis es necesario genotiparlas, para poder asegurar que no se trata de un macho que aún no ha desarrollado su pigmentación característica. Sin embargo, existe constancia de la relevancia que el sexo (tanto las hormonas sexuales como diferencias en comportamiento) tiene en los elementos mediadores de la homeostasis, los cuales pueden ampliamente influenciar la susceptibilidad y resiliencia del proceso de envejecimiento en machos y hembras (véase (Camacho et al., 2009) (Thomas et al., 2019) para ejemplos de la influencia del sexo en envejecimiento). Actualmente, estos experimentos están siendo repetidos con hembras para poder comparar los resultados obtenidos con ambos sexos.

CONCLUSION

El tratamiento del envejecimiento sano y sus enfermedades asociadas es crucial para la mejora de la calidad de vida durante las últimas etapas de nuestra vida. El estudio transcriptómico en combinación con ensayos de accesibilidad genómica en tejidos clave para el envejecimiento, como es el cerebro, suponen herramientas esenciales para entender los cambios que ocurren en las redes génicas involucradas en la regulación de este proceso. Aquí se muestra que hay programas genéticos conservados en las tres principales partes del cerebro, entre los que destacan los procesos de desarrollo y la pérdida de capacidad proliferativa. No obstante, también aparecen relevantes en el proceso de envejecimiento genes asociados a senescencia celular y metabolismo. La identificación de grupos de genes alterados durante envejecimiento da pie al inicio de ensayos funcionales con el fin de validar los análisis bioinformáticos.

METODOS

Diseño experimental

La cría y mantenimiento de *N. furzeri* se llevó a cabo en el animalario del Centro Andaluz de Biología del Desarrollo (CABD) siguiendo el protocolo estándar establecido para esta especie (AU - Dodzian et al., 2018). Para la elaboración de las series temporales de RNA-seq y ATAC-seq se utilizaron cerebros de pez killi extraídos a diferentes puntos temporales. El cerebro completo fue separado en hemisferio izquierdo y hemisferio derecho, usando un hemisferio para RNA-seq y el otro para ATAC-seq, utilizando así ambas técnicas sobre los mismos individuos. Cada hemisferio fue separado en los tres componentes principales del cerebro: telencéfalo, mesencéfalo y romboencéfalo. La serie temporal de RNA-seq posee muestras de los tiempos 8, 12, 16, 20 y 24 semanas después de la salida de diapausa. La serie temporal de ATAC-seq posee muestras de los tiempos 8, 16 y 24 semanas después de la salida de diapausa. Cada punto temporal tiene dos réplicas biológicas.

Genes con expresión diferencial

A partir de los datos de RNA-seq, las lecturas pasaron los filtros de calidad FastQC v0.11.9 y fueron alineadas contra el ensamblaje GRZ (05/2015) (Reichwald et al., 2015) (disponible en <https://nfingb.leibniz-fli.de/>) de *N. furzeri* usando STAR (Dobin et al., 2013) v2.5.3a con el parámetro `--quantMode GeneCounts`. Las lecturas alineadas fueron asignadas a genes utilizando HTSeq (Anders et al., 2015) v0.11.2 con los parámetros `-nonunique all` y `-s no`. Para medir la calidad de las muestras se dibujó una representación PCA eliminando el 10% de los genes menos variables con el paquete de R PCAtools v2.2.0 y las correlaciones de Spearman entre las muestras (Figura suplementaria 1). Los análisis de expresión diferencial se llevaron a cabo usando DESeq2 (Love et al., 2014) v1.30.1. Se usaron dos contrastes de hipótesis: comparación de puntos temporales por parejas usando Wald test y likelihood ratio test (LRT). Para Wald test, un p-valor corregido aplicando la corrección de Benjamini Hochberg (BH) $<0,05$ y un \log_2FC absoluto $>0,75$ fueron utilizados como límites para considerar genes significativos. Para LRT, un p-valor corregido según BH $<0,001$ fue considerado como límite para considerar genes significativos. La figura fue realizada con el paquete de R Pheatmap v1.0.12, usando clustering jerárquico como método de clustering.

Análisis de enriquecimiento en grupos de genes (GSEA)

Para cada tejido, usamos los datos normalizados con el método implementado en DESeq2. Fueron analizados con la aplicación de escritorio GSEA v4.1.0 (Subramanian et al., 2005) los grupos de genes pertenecientes a 'hallmark gene sets' (Liberzon et al., 2015) de la base de datos Molecular Signature Database v7.4. Los parámetros utilizados fueron 1000 permutaciones de los genes, `remap_only` para mapear identificadores y Pearson como métrica para ordenar los genes en cuanto a su correlación con

el tiempo al que fueron recogidas las muestras. Se seleccionaron como significativos aquellos grupos de genes con un p-valor corregido según BH $<0,05$. Se utilizó ggplot2 v3.3.3 (Wickham, 2009) para representar el diagrama de puntos.

Dinámicas de expresión

Los genes con cambios de expresión significativos según LRT fueron clusterizados para identificar patrones de expresión comunes. Se aplicó un algoritmo de clustering fuzzy c-means implementado en el paquete de R Mfuzz v2.50.0 (Futschik M, 2005) y se siguieron los pasos de procesamiento recomendados en el manual. Los conteos de las lecturas fueron transformados a transcritos por millón (TPM) para normalizar las lecturas según la profundidad de secuenciación y la longitud del gen. La longitud de cada gen se calculó con la herramienta GTFtools v0.7.0 (H.-D. Li, 2018), utilizándose la media aritmética de la longitud de todas las isoformas de cada gen. Después se tomó el logaritmo en base 2 para cambiar la escala de los TPMs y se eliminaron los genes que no alcanzaban una media de 1 en la serie temporal. Las réplicas se agregaron usando la media aritmética. A continuación se estandarizaron los valores de expresión de los genes para hacerlos comparables entre diferentes condiciones, teniendo media 0 y desviación típica 1. Finalmente, se ejecutó la normalización por cuantiles implementada en el paquete de R limma v3.46.0 (Ritchie et al., 2015) para hacer comparables las diferentes muestras. Se representó la distribución de los datos para comprobar que no hay sesgos en ninguna muestra usando el paquete de R yarr v0.1.5 (Phillips, 2017) (Figura suplementaria 2).

El número óptimo de clusters se estimó calculando la distancia mínima entre centroides y número de clusters no vacíos (Figura suplementaria 3), considerándose vacío aquel cluster que no contiene ningún gen que alcance un valor de 0,3 para el parámetro *membership*. Se aplicó clustering a cada tejido por separado. Para encontrar clusters con dinámicas similares entre diferentes tejidos se calculó la correlación de Pearson entre los centroides de todos los clusters independientemente del tejido al que pertenecieran. Fueron clusterizados usando kmeans con un número de divisiones arbitrario de 10 (decidido tras el estudio de los grupos en el heatmap). Cada una de esas 10 divisiones sería denominada perfil de expresión.

Se calculó el enriquecimiento en términos GO de procesos biológicos para todos los clusters y los resultados se agruparon por perfil de expresión y por términos GO. Para agrupar los términos GO se hizo uso de expresiones regulares (véase el archivo mfuzz_GO_grafica_puntos.R). Se utilizaron los paquetes de R ggplot2 v3.3.3 y su extensión ggh4x v0.1.2.1 para visualización.

Regiones de la cromatina con accesibilidad diferencial

A partir de los datos de ATAC-seq, las lecturas pasaron los filtros de calidad de FastQC v0.11.9 y fueron alineadas contra el ensamblaje GRZ (05/2015) (disponible en <https://nfingb.leibniz-fli.de/>) de

N. furzeri usando Bowtie2 (Langmead & Salzberg, 2012) con los parámetros `--X 2000 --no-unal --no-mixed`. Así, todas aquellas lecturas que se separen >2 kb son excluidas. Para eliminar artefactos de PCR y duplicados se usó la herramienta rmdup, disponible en Samtools (H. Li et al., 2009). La posición de corte exacta de la transposasa Tn5 se corrigió desplazando las lecturas por +4 o -5 bp en la cadena positiva o negativa, respectivamente. Se seleccionaron aquellas parejas de lecturas que tenían un inserto <130 bp, pues corresponden a lecturas que alinean a regiones no ocupadas por nucleosomas, es decir, regiones abiertas del genoma. A continuación, se generaron los archivos BigWig usando genomeCoverageBed y bedGraphToBigWig de Bedtools (Quinlan & Hall, 2010). Los archivos fueron explorados en el visualizador de genomas UCSC (Kent et al., 2002). El peak calling se calculó con MACS2 (Zhang et al., 2008) usando los parámetros `--nomodel --shift -45 --extsize 100` y el tamaño del genoma, para cada una de las muestras. Los resultados fueron filtrados utilizando las 500.000 RACs con p-valores más bajos. Para la integración de ambas réplicas biológicas se utilizó el método irreproducible discovery rate (IDR) (Q. Li et al., 2011). Se seleccionan las RACs que presentan un alto grado de consistencia entre las réplicas. Para calcular IDR se utilizaron los parámetros `--input-file-type narrowpeak --rank p.value --soft-idr-threshold 0.1`, filtrando las RACs por un valor de IDR global <0,01.

Posteriormente se calculó, para cada tejido, una tabla de conteo de lecturas, conteniendo el número de lecturas que caen en cada pico para cada uno de los puntos de la serie temporal. Para ello, las RACs que pasaron el filtro IDR de toda la serie temporal se unieron en un archivo unificado para después calcular la intersección entre las lecturas de las condiciones por separado con las regiones del archivo unificado. Se utilizó *mergeBed* e *intersectBed* de Bedtools con el parámetro `-c`. Finalmente, se combinaron los filtros por conteos mínimos de DESeq2 (Love et al., 2014) y EdgeR (Robinson et al., 2009) previo a calcular las dRACs (Figura suplementaria 4). Representaciones PCA y correlaciones de Spearman (Figura suplementaria 5) entre las muestras fueron calculadas con las mismas herramientas descritas en el apartado ‘Genes con expresión diferencial’.

Para el análisis diferencial se optó por utilizar el quasi-likelihood F-test implementado en EdgeR v3.32.1, pues los resultados son más robustos que los reportados por LRT cuando se analizan pocas réplicas (según el manual de referencia de EdgeR). Se aplicaron contrastes entre la semana 24 y la 8. Se consideraron como significativos aquellas RACs con un p-valor ajustado según BH <0,05. Las RACs fueron representados utilizando la herramienta deepTools (Ramírez et al., 2016) v3.3.1. Se muestra señal de 1000 bases de distancia en ambas direcciones desde el centro de cada pico.

Descubrimiento de motivos de unión de factores de transcripción

La búsqueda de motivos enriquecidos en los dRACs se llevó a cabo con el programa findMotifsGenome.pl de Homer v4.11.1 usando como total de RACs el archivo de RACs agregados del tejido correspondiente y buscando motivos en la longitud total de las regiones.

También se utilizó el comando `gimmemaelstrom` del paquete de Python `GimmeMotifs` v0.15.2 (Bruse & Heeringen, 2018) para la búsqueda de motivos diferenciales en las dRACs. Se ha dado como input los conteos de lecturas transformados tomando logaritmo y posteriormente centrados por la media, como se especifica en el manual. Finalmente se normaliza para hacer comparaciones entre diferentes experimentos usando normalización por cuantiles. `Gimmemaelstrom` utiliza una base de datos no redundante de motivos en vertebrados para buscar motivos de unión a ADN en las RACs aportadas. `Gimmemaelstrom` combina los resultados de diferentes métodos de regresión y test estadísticos mediante agregación de rangos. Se ha aplicado bayesian ridge regression, multi-task lasso regression, support vector regression y xgboost regression. Los resultados individuales de cada método se ordenan y convierten a z-scores usando la transformación normal inversa. Se muestran los z-scores agregados, calculados usando el método de Stouffer. Un z-score alto está asociado con una elevada señal del motivo en la muestra. Los resultados finales se filtraron eliminando aquellos motivos que, para cada tejido, no alcanzaba un z-score agregado de -3 o +3 en al menos uno de los puntos temporales. Sólo se reportan factores de transcripción experimentalmente validados.

Para priorizar los resultados de `gimmemaelstrom` se emplea `TFEA.ChIP` v1.10.0 (Puente-Santamaria et al., 2019). Esta herramienta estima el enriquecimiento de factores de transcripción en un conjunto de genes diferencialmente expresados. Para ello, busca si hay diferencias significativas en la proporción de los genes con los que un factor de transcripción interacciona, tanto en los genes diferenciales como en el resto. La significancia estadística de la asociación se calcula usando el test exacto de Fisher. Fueron considerados significativos aquellos factores que presentaban un p-valor corregido por BH $<0,05$. Fue configurado para utilizar una base de datos de interacciones entre factores de transcripción y genes proveniente de cientos de experimentos ChIP-seq del proyecto ENCODE (Moore et al., 2020). Dado que los datasets fueron obtenidos de células humanas, los resultados derivados de este método se usarán para validar los resultados de `gimmemaelstrom`.

Integración RNA-seq y ATAC-seq

Las dRACs fueron asignadas a genes usando el comando `closestBed` v2.26.0 de la herramienta `BEDtools` con los parámetros `-D ref -k 2` para buscar los sitios de inicio de transcripción (TSS) más cercanos a cada una de las regiones de cromatina activa. Los TSSs fueron extraídos del ensamblaje GRZ (05/2015). `ClosestBed` se ejecutó para buscar genes más cercanos en ambas direcciones de cada pico (requiere ejecutar `closestBed` con los parámetros `-iu` y `-id` por separado). Se analizará la intersección entre estos genes asociados a dRCAs con los genes cuyo cambio de expresión es significativa según LRT.

Paralelamente, se analizaron los datos de la semana 24 siguiendo una estrategia distinta. Primero, se contaron el número de lecturas de ATAC-seq que solapaban en los promotores de cada gen. La región promotora de un gen se definió como 500 bp en ambas direcciones del TSS. Las lecturas fueron normalizadas con el método normalización por cuantiles y se transformó a escala logarítmica. Los

genes fueron definidos como muy accesibles (HA) si el número de lecturas de ATAC-seq dentro de sus promotores es mayor que el cuantil 0,7 de los datos, y fueron definidos como poco accesibles (LA) si es menor que el cuantil 0,3 de los datos. Los genes fueron definidos como muy expresados (HE) si sus TPMs eran más altos que el cuantil 0,7 de los datos, y fueron definidos como poco expresados (LE) si es menor que el cuantil 0,3 de los datos. Cada gen fue clasificado en uno de los cuatro grupos: LA-LE (poca accesibilidad y expresión), LA-HE (poca accesibilidad y alta expresión), HA-HE (alta accesibilidad y alta expresión) o HA-LE (alta accesibilidad y baja expresión) según sus valores de accesibilidad y expresión. Los genes que no pertenecían a ninguno de estos grupos no fueron usados en los análisis.

Identificación de genes específicos de tejido

Para encontrar genes cuya dinámica de expresión fuera específica de ciertos tejidos se utilizó el test LRT implementado en DESeq2 (Love et al., 2014) v1.30.1 que enfrentaba un modelo completo que incluía los factores tejido, edad e interacción entre tejido y edad contra un modelo nulo que no incluía la interacción entre tejido y edad. El factor edad se configuró como una variable numérica en lugar de como un factor, por lo que genes con p-valores pequeños del test serán aquellos que siguen relaciones lineales distintas. Se consideraron como significativos aquellos RACs con un p-valor ajustado según BH $<0,05$. Esta lista de genes fue interseccionada con la lista de genes que tenían dRACs asociadas.

Enriquecimiento en términos GO

El enriquecimiento en procesos biológicos fue calculado a través del servicio API de PANTHER v16 (Mi et al., 2019) usando arquitectura REST. Se calcula el test de enriquecimiento según el test exacto de Fisher contra la base de datos de genes de pez cebra. Los términos GO enriquecidos con p-valor ajustado según BH $<0,05$ fueron considerados significativos.

Identificación de ortólogos

En primer lugar, los identificadores de genes del ensamblaje GRZ (05/2015) fueron mapeados a identificadores del ensamblaje de *N. furzeri* alojado en Ensembl. Se extrajeron las secuencias de genes con la herramienta getFasta de BEDtools y se preparó la base de datos en local para la ejecución de la herramienta BLAST v2.11.0+ (Camacho et al., 2009) con el comando makeblastdb. Por otro lado, se extrajeron las secuencias de genes del ensamblaje de *N. furzeri* de Ensembl utilizando el paquete de R BiomaRt v2.46.3 (Durinck et al., 2009). Para relacionar los identificadores de Ensembl con los genes de nuestro ensamblaje, se ejecutó blastn y conservamos el mejor resultado acorde a los bits del alineamiento. Se mapearon 24924 identificadores de Ensembl a 19182 identificadores de GRZ (05/2015).

A partir de los identificadores de Ensembl se pueden obtener genes ortólogos en otras especies usando el sistema de consultas implementado en BiomaRt. 21087 y 17800 ortólogos con identificadores HGNC ('gene symbol') fueron encontrados para *D. rerio* y *H. sapiens* a partir de los 24924 identificadores de *N. furzeri*.

Análisis de asociación con el envejecimiento

Se descargaron las bases de datos GenAge Human, DrugAge y CellAge (Tacutu et al., 2018). Se usaron los identificadores HGNC de los ortólogos en humano para utilizar estas bases de datos en *N. furzeri*. Para obtener los genes que son diana de los medicamentos en la base de datos DrugAge, se usaron los nombres de los medicamentos presentes en la base de datos para mapear los genes con los que interaccionan usando la base de datos DGIdb 4.0 (Freshour et al., 2021). Esta última base de datos fue filtrada para mantener interacciones gen-medicamento con altos valores de confianza. No fueron tenidas en consideración interacciones no consideradas significativas ni interacciones con un score inferior a 3. Se calcula el test exacto de Fisher para establecer si existe relación significativa entre un grupo de genes y la pertenencia a las diferentes bases de datos.

Inyección de embriones

Con propósito del aprendizaje de inyección de embriones de *N. furzeri*, practiqué el procedimiento en pez cebra. El primer paso consiste en la recolección de embriones, para lo cual machos y hembras son separados (reconociendo las hembras por la distintiva salida de la cloaca entre las aletas pélvicas y caudal) el día previo al procedimiento. Esto facilita el apareamiento cuando se ponen juntos al día siguiente justo tras el encendido de luces en el animalario. Los embriones se recolectan en el mismo medio acuoso en el que viven los adultos al que se añade un fungicida (azul de metileno). El estadio adecuado para inyectar constructos genéticos y producir individuos transgénicos sin mosaicismo es en embrión de una célula. Por tanto, las inyecciones deben producirse a los pocos minutos tras la fertilización. Para inyectar, se utiliza un microinyector acoplado a manipulador (Figura 7a) que aplica pequeños pulsos de presión para introducir el vector de transformación en el embrión. Se inyectó el vector ZED (del inglés, Zebrafish Enhancer Detector) (Bessa et al., 2009) portando el gen reportero GFP, el promotor mínimo GATA2 y clonada la secuencia activadora de la transcripción de POU3F4 (Naranjo et al., 2010), específica de hinbrain y espina dorsal en pez cebra. La integración del vector en el genoma del embrión producirá la expresión de la GFP como consecuencia del montaje de los mecanismos de inicio de transcripción debido a la secuencia enhancer. 24 horas tras la fertilización puede verse expresión de GFP (Figura 7b).

Código disponible en <https://github.com/adeben1/killifish-aging-brain>

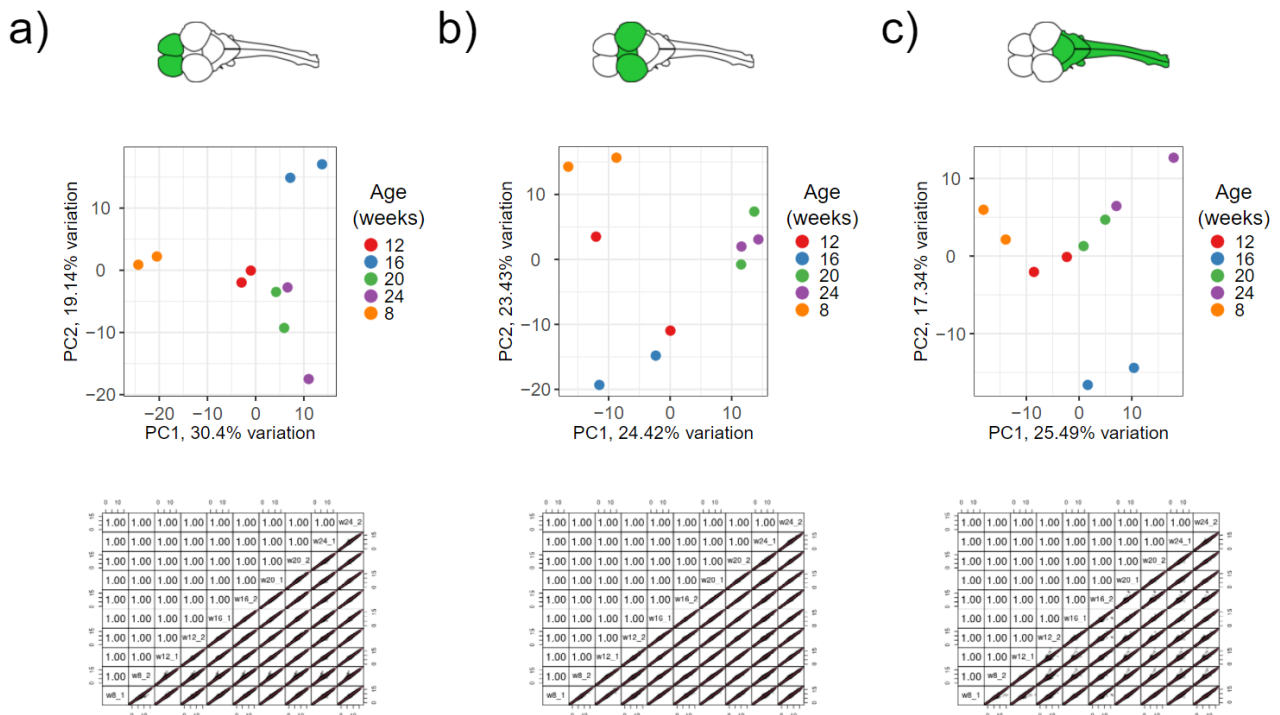


Figura suplementaria 1. Control de calidad de las muestras de RNA-seq. Se usaron los datos normalizados con el método implementado en DESeq2. a) Análisis de componentes principales, habiéndose eliminado el 10% de los genes con menor variabilidad. b) Correlación de Spearman por parejas. De izquierda a derecha: telencéfalo, mesencéfalo y romboencéfalo.

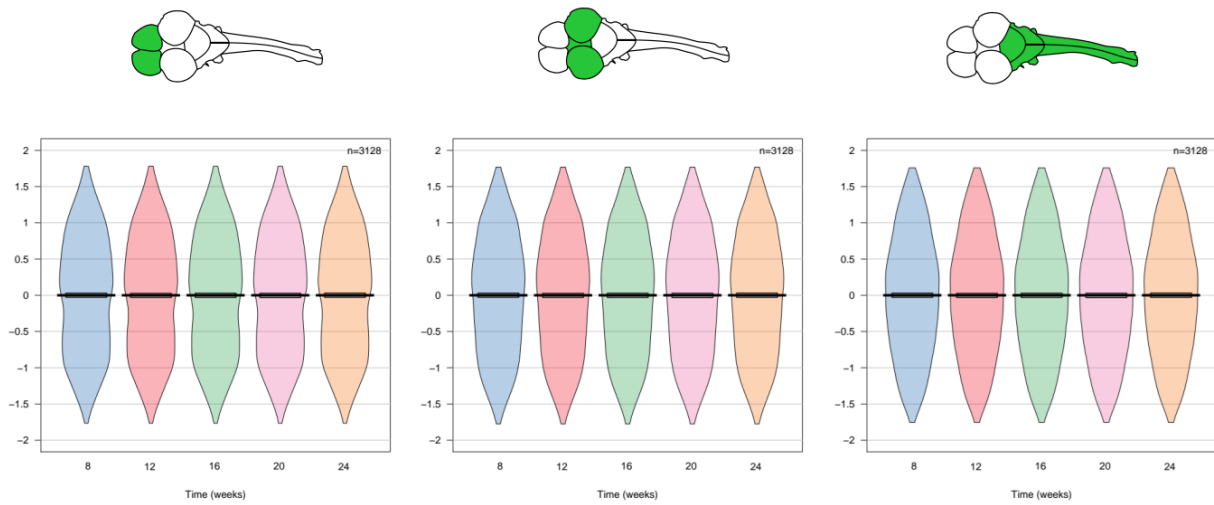


Figura suplementaria 2. Distribución del conteo de lecturas tras el preprocesamiento. Cada distribución presenta la media como una línea negra y los intervalos de confianza al 95%. De izquierda a derecha: telencéfalo, mesencéfalo y romboencéfalo.

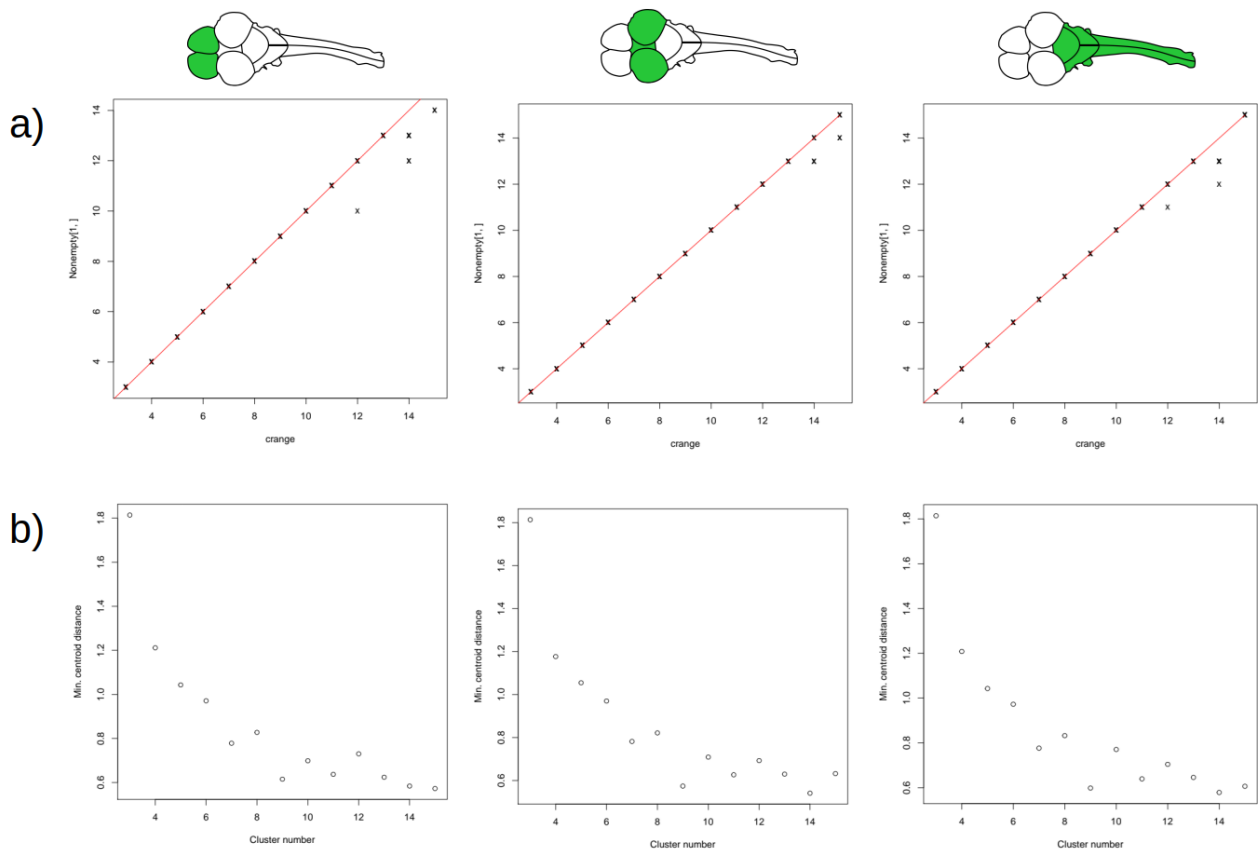


Figura suplementaria 3. Número de clusters óptimo para cada tejido. Se usaron los criterios: (a) número de clusters vacíos y (b) distancia mínima entre centroides. Se aprecia que aparecen clusters vacíos cuando los datos se dividen en 12 clusters o más y que aparece un cambio en la pendiente de la distancia mínima entre clusters cuando los datos se dividen en 9 clusters. Ambos métodos coinciden en que 9 clusters para cada tejido es una división adecuada. De izquierda a derecha: telencéfalo, mesencéfalo y romboencéfalo.

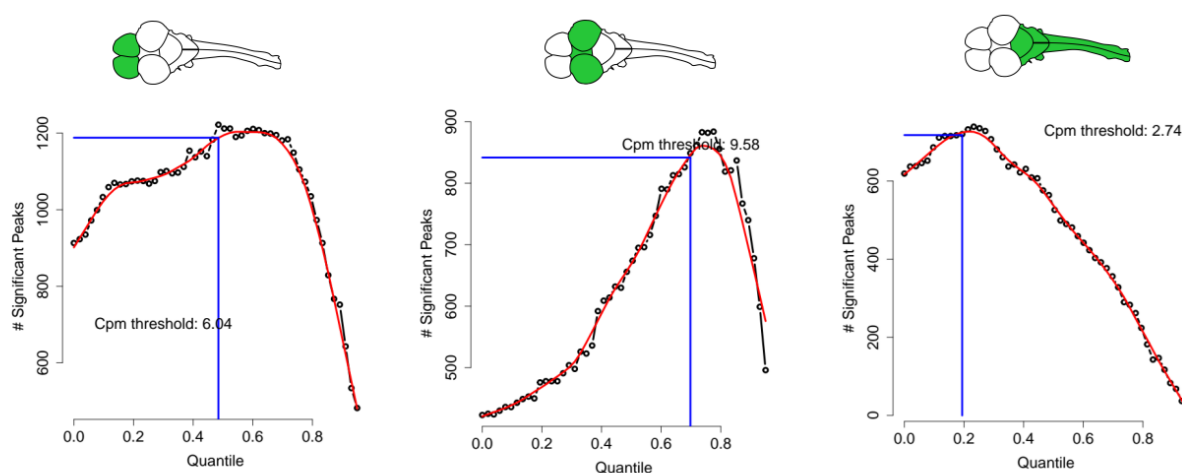


Figura suplementaria 4. Límite por bajos conteos óptimo para diferenciales. Se combinan los métodos de filtrado por bajos conteos implementados en DESeq2 y EdgeR. Las RACs incluidas en los diferenciales debían tener en al menos 2 muestras con más de 2 conteos por millón (CPM). Se usaron cuantiles en lugar de CPMs para ser independiente de la profundidad de secuenciación. Se realizaron 50 diferenciales usando EdgeR. De cada una se recoge el número de dRACs (p-valor según BH $<0,05$) y se ajusta una curva a los puntos. El límite óptimo por conteos mínimos será el máximo de la curva menos la desviación estándar entre la línea ajustada y los valores originales. De izquierda a derecha: telencéfalo, mesencéfalo y romboencéfalo.

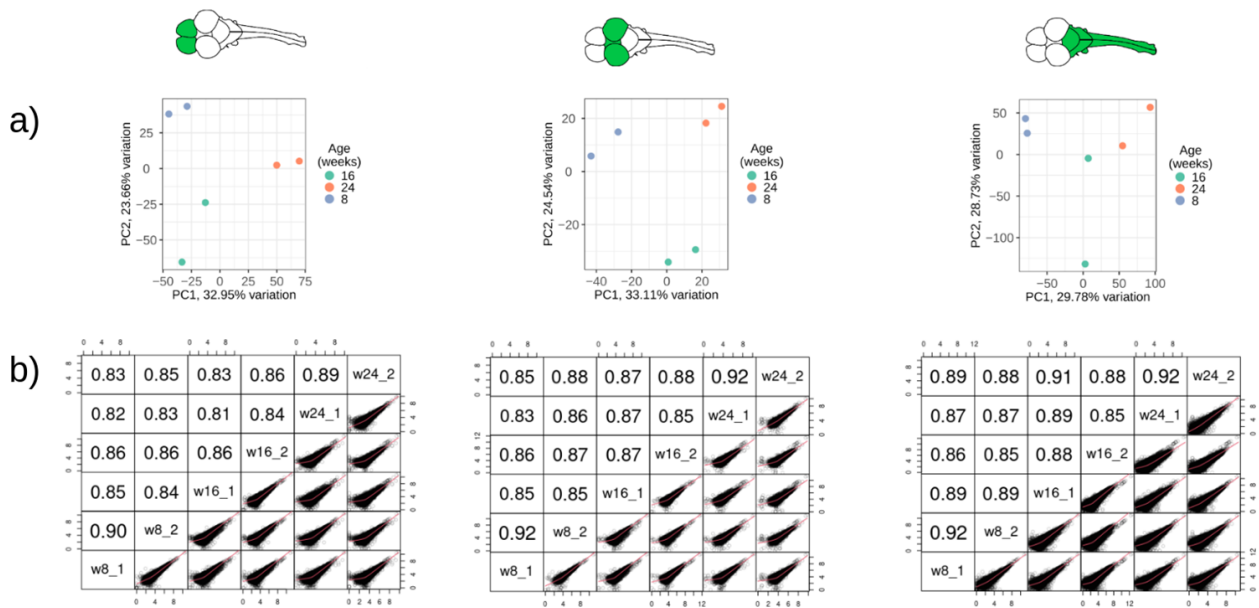


Figura suplementaria 5. Control de calidad de las muestras de ATAC-seq. Se usaron los datos normalizados con el método implementado en DESeq2. a) Análisis de componentes principales, habiéndose eliminado el 10% de los genes con menor variabilidad. b) Correlación de Spearman por parejas.

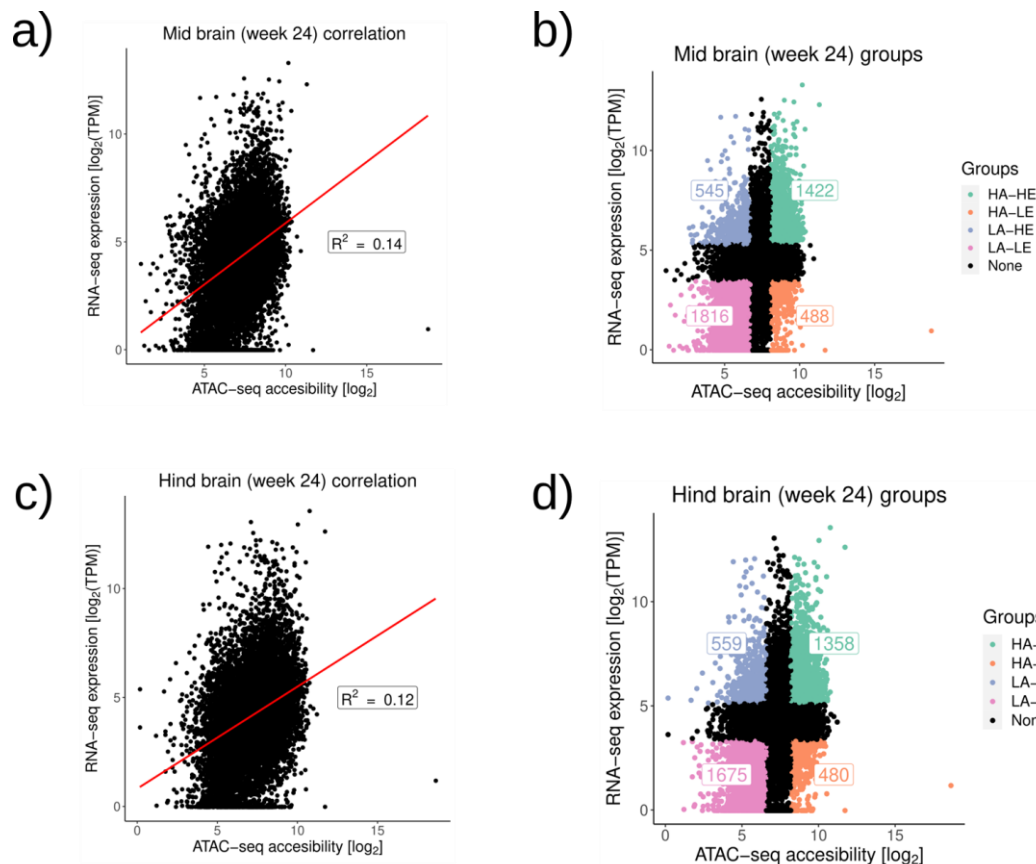


Figura suplementaria 6. Accesibilidad de los promotores en la semana 24 en mesencéfalo y romboencéfalo. a) R^2 en mesencéfalo. b) Grupos de genes en mesencéfalo. c) R^2 en romboencéfalo. d) Grupos de genes en romboencéfalo.

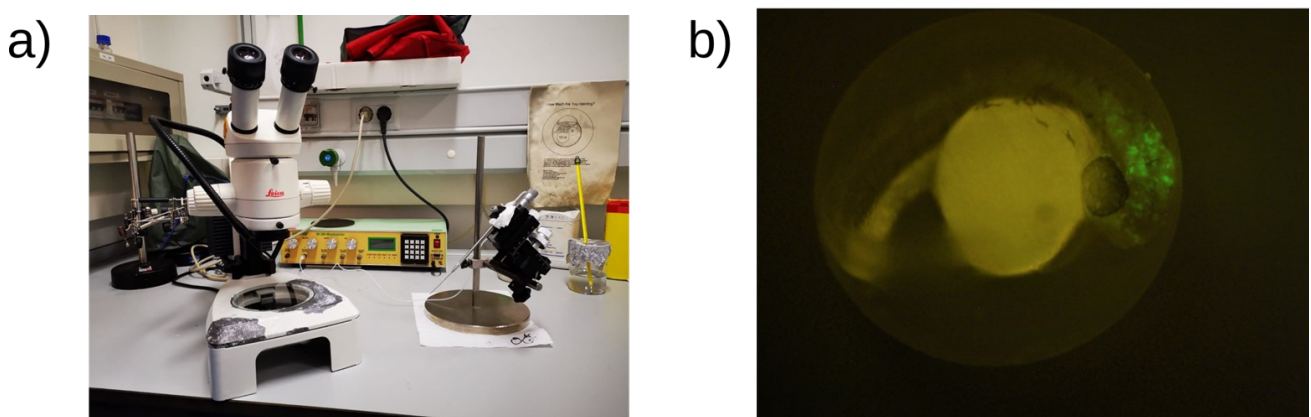


Figura suplementaria 7. Inyección de embriones de pez cebra. a) Instrumentos utilizados para inyectar. A la izquierda: lentes; al fondo: manipulador de la presión; a la derecha: microinyector. b) Embrión transgénico de pez cebra 24 horas tras la fertilización. Se ve expresión de GFP en romboencéfalo.

BIBLIOGRAFÍA

- Ali, A. A. H., Tundo-Lavalle, F., Hassan, S. A., Pfeffer, M., Stahr, A., & von Gall, C. (2020). Impact of Targeted Deletion of the Circadian Clock Gene *Bmal1* in Excitatory Telencéfalo Neurons on Adult Neurogenesis and Olfactory Function. *International Journal of Molecular Sciences*, 21(4). <https://doi.org/10.3390/ijms21041394>
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- AU - Dodzian, J., AU - Kean, S., AU - Seidel, J., & AU - Valenzano, D. R. (2018). A Protocol for Laboratory Housing of Turquoise Killifish (*Nothobranchius furzeri*). *JoVE*, 134, e57073. <https://doi.org/doi:10.3791/57073>
- Autuori, E., Sedlak, P., Xu, L., C Ridder, M., Tedoldi, A., & Sah, P. (2019). rSK1 in Rat Neurons: A Controller of Membrane rSK2? *Frontiers in Neural Circuits*, 13, 21. <https://doi.org/10.3389/fncir.2019.00021>
- Bessa, J., Tena, J. J., de la Calle-Mustienes, E., Fernández-Miñán, A., Naranjo, S., Fernández, A., Montoliu, L., Akalin, A., Lenhard, B., Casares, F., & Gómez-Skarmeta, J. L. (2009). Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, 238(9), 2409–2417. <https://doi.org/10.1002/dvdy.22051>
- Bruse, N., & Heeringen, S. J. van. (2018). GimmeMotifs: an analysis framework for transcription factor motif analysis. *BioRxiv*, 474403. <https://doi.org/10.1101/474403>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 1–9. <https://doi.org/10.1186/1471-2105-10-421>
- Canivet, C., Rhéaume, C., Lebel, M., Piret, J., Gosselin, J., & Boivin, G. (2018). Both IRF3 and especially IRF7 play a key role to orchestrate an effective cerebral inflammatory response in a mouse model of herpes simplex virus encephalitis. *Journal of Neurovirology*, 24(6), 761–768. <https://doi.org/10.1007/s13365-018-0666-9>
- Cellerino, A., Valenzano, D. R., & Reichard, M. (2016). From the bush to the bench: The annual *Nothobranchius* fishes as a new model system in biology. *Biological Reviews*, 91(2), 511–533. <https://doi.org/10.1111/brv.12183>
- Choi, S.-H., Koh, D.-I., Cho, S.-Y., Kim, M.-K., Kim, K.-S., & Hur, M.-W. (2019). Temporal and differential regulation of KAISO-controlled transcription by phosphorylated and acetylated p53

- highlights a crucial regulatory role of apoptosis. *The Journal of Biological Chemistry*, 294(35), 12957–12974. <https://doi.org/10.1074/jbc.RA119.008100>
- Conceição, R., Evans, R. S., Pearson, C. S., Hänzi, B., Osborne, A., Deshpande, S. S., Martin, K. R., & Barber, A. C. (2019). Expression of Developmentally Important Axon Guidance Cues in the Adult Optic Chiasm. *Investigative Ophthalmology & Visual Science*, 60(14), 4727–4739. <https://doi.org/10.1167/iovs.19-26732>
- da Costa, J. P., Vitorino, R., Silva, G. M., Vogel, C., Duarte, A. C., & Rocha-Santos, T. (2016). A synopsis on aging—Theories, mechanisms and future prospects. *Ageing Research Reviews*, 29, 90–112. <https://doi.org/10.1016/j.arr.2016.06.005>
- Dang, C. V. (1999). c-Myc Target Genes Involved in Cell Growth, Apoptosis, and Metabolism. *Molecular and Cellular Biology*, 19(1), 1 LP – 11. <https://doi.org/10.1128/MCB.19.1.1>
- Daniel S. Evans, P. K., & , Wen-Chi Hsueh, and L. K. (2011). TOR signaling never gets old: Aging, longevity and TORC1 activity. *Ageing Res Rev*, 10(2), 225–237. <https://doi.org/10.1016/j.arr.2010.04.001>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4(8), 1184–1191. <https://doi.org/10.1038/nprot.2009.97>
- Franke, M., De la Calle-Mustienes, E., Neto, A., Acemel, R. D., Tena, J. J., Santos-Pereira, J. M., & Gómez-Skarmeta, J. L. (2020). CTCF knockout in zebrafish induces alterations in regulatory landscapes and developmental gene expression. *BioRxiv*, 2020.09.08.282707. <https://doi.org/10.1101/2020.09.08.282707>
- Freshour, S. L., Kiwala, S., Cotto, K. C., Coffman, A. C., McMichael, J. F., Song, J. J., Griffith, M., Griffith, O. L., & Wagner, A. H. (2021). Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsourced efforts. *Nucleic Acids Research*, 49(D1), D1144–D1151. <https://doi.org/10.1093/nar/gkaa1084>
- Furigo, I. C., Teixeira, P. D. S., Quaresma, P. G. F., Mansano, N. S., Frazão, R., & Donato, J. (2020). STAT5 ablation in AgRP neurons increases female adiposity and blunts food restriction adaptations. *Journal of Molecular Endocrinology*, 64(1), 13–27. <https://doi.org/10.1530/JME-19-0158>

- Futschik M, C. B. (2005). Noise robust clustering of gene expression time-course data. *Journal of Bioinformatics and Computational Biology*, 965–988. <http://mfuzz.sysbiolab.eu>
- Grabowska, W., Sikora, E., & Bielak-Zmijewska, A. (2017). Sirtuins, a promising target in slowing down the ageing process. *Biogerontology*, 18(4), 447–476. <https://doi.org/10.1007/s10522-017-9685-9>
- Harel, I., Benayoun, B. A., Machado, B., Singh, P. P., Hu, C. K., Pech, M. F., Valenzano, D. R., Zhang, E., Sharp, S. C., Artandi, S. E., & Brunet, A. (2015). A platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate. *Cell*, 160(5), 1013–1026. <https://doi.org/10.1016/j.cell.2015.01.038>
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, 12(6), 996–1006. <https://doi.org/10.1101/gr.229102>
- Kenyon, C. J. (2010). The genetics of ageing. *Nature*, 464(7288), 504–512. <https://doi.org/10.1038/nature08980>
- Kenyon, C., Jean, C., Gensch, E., Adam, R., & Ramon, T. (1993). A *C.elegans* mutant that twice as long as wild type. *Nature*, 366(6454), 461–464.
- Kondratov, R. V, Kondratova, A. A., Gorbacheva, V. Y., Vykhovanets, O. V, & Antoch, M. P. (2006). Early aging and age-related pathologies in mice deficient in BMAL1, the core component of the circadian clock. *Genes & Development*, 20(14), 1868–1873. <https://doi.org/10.1101/gad.1432206>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lanzillotta, A., Porrini, V., Bellucci, A., Benarese, M., Branca, C., Parrella, E., Spano, P. F., & Pizzi, M. (2015). NF-κB in Innate Neuroprotection and Age-Related Neurodegenerative Diseases. *Frontiers in Neurology*, 6, 98. <https://doi.org/10.3389/fneur.2015.00098>
- Lee, J. Y., Lake, R. J., Kirk, J., Bohr, V. A., Fan, H.-Y., & Hohng, S. (2017). NAP1L1 accelerates activation and decreases pausing to enhance nucleosome remodeling by CSB. *Nucleic Acids Research*, 45(8), 4696–4707. <https://doi.org/10.1093/nar/gkx188>
- Li, H.-D. (2018). GTFtools: a Python package for analyzing various modes of gene models. *BioRxiv*, 263517. <https://doi.org/10.1101/263517>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., &

- Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, Q., Brown, J. B., Huang, H., & Bickel, P. J. (2011). Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics*, 5(3), 1752–1779. <https://doi.org/10.1214/11-AOAS466>
- Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Systems*, 1(6), 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>
- Liscovitch, N., & Chechik, G. (2013). Specialization of Gene Expression during Mouse Brain Development. *PLOS Computational Biology*, 9(9), e1003185. <https://doi.org/10.1371/journal.pcbi.1003185>
- Liu, P., Gao, Q., Guan, L., Sheng, W., Hu, Y., Gao, T., Jiang, J., Xu, Y., Qiao, H., Xue, X., Liu, S., & Li, T. (2020). Atorvastatin Attenuates Isoflurane-Induced Activation of ROS-p38MAPK/ATF2 Pathway, Neuronal Degeneration, and Cognitive Impairment of the Aged Mice. *Frontiers in Aging Neuroscience*, 12, 620946. <https://doi.org/10.3389/fnagi.2020.620946>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 1–21. <https://doi.org/10.1186/s13059-014-0550-8>
- Lui, J. C., Chen, W., Barnes, K. M., & Baron, J. (2010). Changes in gene expression associated with aging commonly originate during juvenile growth. *Mechanisms of Ageing and Development*, 131(10), 641–649. <https://doi.org/10.1016/j.mad.2010.08.010>
- Mattson, M. P., & Arumugam, T. V. (2018). Hallmarks of Brain Aging: Adaptive and Pathological Modification by Metabolic States. *Cell Metabolism*, 27(6), 1176–1199. <https://doi.org/10.1016/j.cmet.2018.05.011>
- Melzer, D., Pilling, L. C., & Ferrucci, L. (2020). The genetics of human ageing. *Nature Reviews Genetics*, 21(2), 88–101. <https://doi.org/10.1038/s41576-019-0183-6>
- Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, 47(D1), D419–D426. <https://doi.org/10.1093/nar/gky1038>
- Mitchell, S. J., Scheibye-Knudsen, M., Longo, D. L., & De Cabo, R. (2015). Animal models of aging research: Implications for human aging and age-related diseases. *Annual Review of Animal Biosciences*, 3, 283–303. <https://doi.org/10.1146/annurev-animal-022114-110829>

- Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shores, N., Adrian, J., Kawli, T., Davis, C. A., Dobin, A., Kaul, R., Halow, J., Van Nostrand, E. L., Freese, P., Gorkin, D. U., Shen, Y., He, Y., Mackiewicz, M., Pauli-Behn, F., Williams, B. A., ... Consortium, T. E. P. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, 583(7818), 699–710. <https://doi.org/10.1038/s41586-020-2493-4>
- Naranjo, S., Voeselek, K., de la Calle-Mustienes, E., Robert-Moreno, A., Kokotas, H., Grigoriadou, M., Economides, J., Van Camp, G., Hilgert, N., Moreno, F., Alsina, B., Petersen, M. B., Kremer, H., & Gómez-Skarmeta, J. L. (2010). Multiple enhancers located in a 1-Mb region upstream of POU3F4 promote expression during inner ear development and may be required for hearing. *Human Genetics*, 128(4), 411–419. <https://doi.org/10.1007/s00439-010-0864-x>
- Newgard, C. B., & Sharpless, N. E. (2013). Coming of age: Molecular drivers of aging and therapeutic opportunities. *Journal of Clinical Investigation*, 123(3), 946–950. <https://doi.org/10.1172/JCI68833>
- Phillips, N. (2017). *yarr: A Companion to the e-Book “YaRrr!: The Pirate’s Guide to R.”* <https://cran.r-project.org/package=yarr>
- Puente-Santamaria, L., Wasserman, W. W., & del Peso, L. (2019). TFEA.ChIP: a tool kit for transcription factor binding site enrichment analysis capitalizing on ChIP-seq datasets. *Bioinformatics*, 35(24), 5339–5340. <https://doi.org/10.1093/bioinformatics/btz573>
- Pugacheva, E. M., Rivero-Hinojosa, S., Espinoza, C. A., Méndez-Catalá, C. F., Kang, S., Suzuki, T., Kosaka-Suzuki, N., Robinson, S., Nagarajan, V., Ye, Z., Boukaba, A., Rasko, J. E. J., Strunniakov, A. V., Loukinov, D., Ren, B., & Lobanenko, V. V. (2015). Comparative analyses of CTCF and BORIS occupancies uncover two distinct classes of CTCF binding genomic regions. *Genome Biology*, 16(1), 161. <https://doi.org/10.1186/s13059-015-0736-8>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1), W160–W165. <https://doi.org/10.1093/nar/gkw257>
- Reichwald, K., Petzold, A., Koch, P., Downie, B. R., Hartmann, N., Pietsch, S., Baumgart, M., Chalopin, D., Felder, M., Bens, M., Sahm, A., Szafranski, K., Taudien, S., Groth, M., Arisi, I., Weise, A., Bhatt, S. S., Sharma, V., Kraus, J. M., ... Platzer, M. (2015). Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. *Cell*, 163(6), 1527–1538. <https://doi.org/10.1016/j.cell.2015.10.071>

- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47–e47. <https://doi.org/10.1093/nar/gkv007>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Somel, M., Guo, S., Fu, N., Yan, Z., Hu, H. Y., Xu, Y., Yuan, Y., Ning, Z., Hu, Y., Menzel, C., Hu, H., Lachmann, M., Zeng, R., Chen, W., & Khaitovich, P. (2010). MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain. *Genome Research*, 20(9), 1207–1218. <https://doi.org/10.1101/gr.106849.110>
- Suberbielle, E., Djukic, B., Evans, M., Kim, D. H., Taneja, P., Wang, X., Finucane, M., Knox, J., Ho, K., Devidze, N., Masliah, E., & Mucke, L. (2015). DNA repair factor BRCA1 depletion occurs in Alzheimer brains and impairs cognitive function in mice. *Nature Communications*, 6, 8897. <https://doi.org/10.1038/ncomms9897>
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), 15545 LP – 15550. <https://doi.org/10.1073/pnas.0506580102>
- Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V. E., & de Magalhães, J. P. (2018). Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Research*, 46(D1), D1083–D1090. <https://doi.org/10.1093/nar/gkx1042>
- Terzibasi, E., Valenzano, D. R., Benedetti, M., Roncaglia, P., Cattaneo, A., Domenici, L., & Cellerino, A. (2008). Large differences in aging phenotype between strains of the short-lived annual fish *Nothobranchius furzeri*. *PLoS ONE*, 3(12). <https://doi.org/10.1371/journal.pone.0003866>
- Thomas, N., Gurvich, C., & Kulkarni, J. (2019). Sex Differences in Aging and Associated Biomarkers. *Advances in Experimental Medicine and Biology*, 1178, 57–76. https://doi.org/10.1007/978-3-030-25650-0_4
- Truty, M. J., Lomberk, G., Fernandez-Zapico, M. E., & Urrutia, R. (2009). Silencing of the transforming growth factor-beta (TGFbeta) receptor II by Kruppel-like factor 14 underscores the importance of a negative feedback mechanism in TGFbeta signaling. *The Journal of Biological Chemistry*, 284(10), 6291–6300. <https://doi.org/10.1074/jbc.M807791200>

- Van den Berge, K., Hembach, K. M., Soneson, C., Tiberi, S., Clement, L., Love, M. I., Patro, R., & Robinson, M. D. (2019). RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis. *Annual Review of Biomedical Data Science*, 2(1), 139–173. <https://doi.org/10.1146/annurev-biodatasci-072018-021255>
- Vitale, G., Pellegrino, G., Vollery, M., & Hofland, L. J. (2019). ROLE of IGF-1 system in the modulation of longevity: Controversies and new insights from a centenarians' perspective. *Frontiers in Endocrinology*, 10(FEB), 1–11. <https://doi.org/10.3389/fendo.2019.00027>
- Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T. W., Greven, M. C., Pierce, B. G., Dong, X., Kundaje, A., Cheng, Y., Rando, O. J., Birney, E., Myers, R. M., Noble, W. S., Snyder, M., & Weng, Z. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Research*, 22(9), 1798–1812. <https://doi.org/10.1101/gr.139105.112>
- Weiss, A., & Attisano, L. (2013). The TGFbeta superfamily signaling pathway. *Wiley Interdisciplinary Reviews: Developmental Biology*, 2(1), 47–63. <https://doi.org/10.1002/wdev.86>
- Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*.
- Xiao, T., Li, X., & Felsenfeld, G. (2021). The Myc-associated zinc finger protein (MAZ) works together with CTCF to control cohesin positioning and genome organization. *Proceedings of the National Academy of Sciences of the United States of America*, 118(7), 1–10. <https://doi.org/10.1073/pnas.2023127118>
- Yamagata, M., Duan, X., & Sanes, J. R. (2018). Cadherins Interact With Synaptic Organizers to Promote Synaptic Differentiation. *Frontiers in Molecular Neuroscience*, 11, 142. <https://doi.org/10.3389/fnmol.2018.00142>
- Yan, F., Powell, D. R., Curtis, D. J., & Wong, N. C. (2020). From reads to insight: A hitchhiker's guide to ATAC-seq data analysis. *Genome Biology*, 21(1), 1–16. <https://doi.org/10.1186/s13059-020-1929-3>
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Shirley, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9). <https://doi.org/10.1186/gb-2008-9-9-r137>